

---

# **Snorkel Documentation**

*Release 0.6.0*

**Alex Ratner, Stephen Bach, Henry Ehrenberg**

**Jul 25, 2017**



---

## Contents

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Contexts</b>   | <b>3</b>  |
| 1.1      | Core Data Models . . . . .                                | 3         |
| 1.2      | Core Objects for Preprocessing and Loading . . . . .      | 4         |
| <b>2</b> | <b>Candidates</b>   | <b>7</b>  |
| 2.1      | Core Data Models . . . . .                                | 7         |
| 2.2      | Core Objects for Candidate Extraction . . . . .           | 8         |
| <b>3</b> | <b>Annotations</b>  | <b>11</b> |
| 3.1      | Core Data Models . . . . .                                | 11        |
| 3.2      | Core Objects for Annotations (Features, Labels) . . . . . | 12        |
| <b>4</b> | <b>Learning</b>   | <b>15</b> |
| 4.1      | Base Classifier Class . . . . .                           | 15        |
| 4.2      | Generative Model . . . . .                                | 16        |
| 4.3      | Discriminative Models . . . . .                           | 19        |
| 4.4      | Learning Utilities . . . . .                              | 29        |
| <b>5</b> | <b>Etc: Viewing and Annotating Data, Writing LFs</b>      | <b>33</b> |
| 5.1      | Using the Viewer to Inspect and Annotate Data . . . . .   | 33        |
| 5.2      | Helpers for Writing Labeling Functions . . . . .          | 33        |
| 5.3      | Helpers for Loading External Annotations . . . . .        | 34        |
|          | <b>Python Module Index</b>                                | <b>35</b> |





**snorkel**



Preprocessed input data is represented in Snorkel as a hierarchy of *Context* subclass objects. For example, as is currently default for text: Corpus -> Document -> Sentence -> Span.

## Core Data Models

**class** `snorkel.models.context.Context` (\*\*kwargs)

A piece of content from which Candidates are composed.

**class** `snorkel.models.context.Document` (\*\*kwargs)

A root Context.

**class** `snorkel.models.context.Sentence` (\*\*kwargs)

A sentence Context in a Document.

**class** `snorkel.models.context.Span` (\*\*kwargs)

A span of characters, identified by Context id and character-index start, end (inclusive).

`char_offsets` are **relative to the Context start**

**class** `snorkel.models.context.TemporaryContext`

A context which does not incur the overhead of a proper ORM-based Context object. The `TemporaryContext` class is specifically for the candidate extraction process, during which a `CandidateSpace` object will generate many `TemporaryContexts`, which will then be filtered by `Matchers` prior to materialization of `Candidates` and constituent `Context` objects.

Every `Context` object has a corresponding `TemporaryContext` object from which it inherits.

A `TemporaryContext` must have specified equality / set membership semantics, a `stable_id` for checking uniqueness against the database, and a `promote()` method which returns a corresponding `Context` object.

**class** `snorkel.models.context.TemporarySpan` (*sentence*, *char\_start*, *char\_end*, *meta=None*)

The `TemporaryContext` version of `Span`

**char\_to\_word\_index** (*ci*)

Given a character-level index (offset), return the index of the **word this char is in**

**get\_attrib\_span** (*a*, *sep*=' ')

Get the span of sentence attribute *\_a\_* over the range defined by *word\_offset*, *n*

**get\_attrib\_tokens** (*a*='words')

Get the tokens of sentence attribute *\_a\_* over the range defined by *word\_offset*, *n*

**word\_to\_char\_index** (*wi*)

Given a word-level index, return the character-level index (offset) of the word's start

`snorkel.models.context.construct_stable_id` (*parent\_context*, *polymorphic\_type*,  
*relative\_char\_offset\_start*, *relative\_char\_offset\_end*)

Contract a stable ID for a Context given its parent and its character offsets relative to the parent

`snorkel.models.context.split_stable_id` (*stable\_id*)

**Split stable id, returning:**

- Document (root) stable ID
- Context polymorphic type
- Character offset start, end *relative to document start*

Returns tuple of four values.

## Core Objects for Preprocessing and Loading

`class snorkel.parser.doc_preprocessors.CSVPathsPreprocessor` (*path*,  
*parser\_factory*=<class  
'snorkel.parser.doc\_preprocessors.TextDocPreprocessor',  
*column*=None, *delim*='',  
, \**args*, \*\**kwargs*)

This *DocumentPreprocessor* treats inputs file as index of paths to actual documents; each line in the input file contains a path to a document.

**Defaults and Customization:**

- The input file is treated as a simple text file having one path per file. However, if the input is a CSV file, a pair of *column* and *delim* parameters may be used to retrieve the desired value as reference path.
- The referenced documents are treated as text document and hence parsed using *TextDocPreprocessor*. However, if the referenced files are complex, an advanced parser may be used by specifying *parser\_factory* parameter to constructor.

`class snorkel.parser.doc_preprocessors.DocPreprocessor` (*path*, *encoding*='utf-8',  
*max\_docs*=inf)

Processes a file or directory of files into a set of Document objects.

**Parameters**

- **encoding** – file encoding to use, default='utf-8'
- **path** – filesystem path to file or directory to parse
- **max\_docs** – the maximum number of Documents to produce, default=float('inf')

**generate** ()

Parses a file or directory of files into a set of Document objects.



**class** snorkel.parser.doc\_preprocessors.**HTMLDocPreprocessor** (*path*, *encoding='utf-8'*,  
*max\_docs=inf*)

Simple parsing of raw HTML files, assuming one document per file

**class** snorkel.parser.doc\_preprocessors.**TSVDocPreprocessor** (*path*, *encoding='utf-8'*,  
*max\_docs=inf*)

Simple parsing of TSV file with one (doc\_name <tab> doc\_text) per line

**class** snorkel.parser.doc\_preprocessors.**TextDocPreprocessor** (*path*, *encoding='utf-8'*,  
*max\_docs=inf*)

Simple parsing of raw text files, assuming one document per file

**class** snorkel.parser.doc\_preprocessors.**TikaPreprocessor** (*path*, *encoding='utf-8'*,  
*max\_docs=inf*)

This preprocessor use [Apache Tika](#) parser to retrieve text content from complex file types such as DOCX, HTML and PDFs.

Documentation for customizing Tika is [here](#)

Example:

```
!find pdf_dir -name *.pdf > input.csv # list of files
from snorkel.parser import (
    TikaPreprocessor, CSVPathsPreprocessor, CorpusParser
)
CorpusParser().apply(
    CSVPathsPreprocessor('input.csv', parser_factory=TikaPreprocessor)
)
```

**class** snorkel.parser.doc\_preprocessors.**XMLMultiDocPreprocessor** (*path*,  
*doc='//document'*,  
*text='./text/text()'*,  
*id='./id/text()'*,  
*keep\_xml\_tree=False*)

Parse an XML file *\_which* contains multiple **documents** into a set of Document objects.

Use XPath queries to specify a *\_document\_* object, and then for each document, a set of *\_text\_* sections and an *\_id\_*.

**Note: Include the full document XML etree in the attribs dict with keep\_xml\_tree=True**

**class** snorkel.parser.parser.**ParserConnection** (*parser*)  
Default connection object assumes local parser object

**class** snorkel.parser.parser.**URLParserConnection** (*parser*, *retries=5*)  
URL parser connection

**parse** (*document*, *text*)

Return parse generator :param document: :param text: :return:

**post** (*url*, *data*, *allow\_redirects=True*)

#### Parameters

- **url** –
- **data** –
- **allow\_redirects** –
- **timeout** –

#### Returns



In order to apply machine learning—i.e., in this case, a classifier—to information extraction problems, we need to have a base set of objects that are being classified. In Snorkel, these are the *Candidate* subclasses, which are defined over *Context* arguments, and represent *potential* mentions to extract. We use *Matcher* operators to extract a set of *Candidate* objects from the input data.

## Core Data Models

**class** snorkel.models.candidate.**Candidate** (\*\*kwargs)

An abstract candidate relation.

New relation types should be defined by calling `candidate_subclass()`, **not** subclassing this class directly.

**get\_cids** ()

Get a tuple of the canonical IDs (CIDs) of the contexts making up this candidate

**get\_contexts** ()

Get a tuple of the constituent contexts making up this candidate

**class** snorkel.models.candidate.**Marginal** (\*\*kwargs)

A marginal probability corresponding to a (Candidate, value) pair.

Represents:

$P(\text{candidate} = \text{value}) = \text{probability}$

@training: If True, this is a training marginal; otherwise is end prediction

snorkel.models.candidate.**candidate\_subclass** (class\_name, args, table\_name=None, cardinality=None, values=None)

Creates and returns a Candidate subclass with provided argument names, which are Context type. Creates the table in DB if does not exist yet.

Import using:

```
from snorkel.models import candidate_subclass
```

### Parameters

- **class\_name** – The name of the class, should be “camel case” e.g. NewCandidate
- **args** – A list of names of constituent arguments, which refer to the Contexts—representing mentions—that comprise the candidate
- **table\_name** – The name of the corresponding table in DB; if not provided, is converted from camel case by default, e.g. new\_candidate
- **cardinality** – The cardinality of the variable corresponding to the Candidate. By default is 2 i.e. is a binary value, e.g. is or is not a true mention.

## Core Objects for Candidate Extraction

```
class snorkel.candidates.CandidateExtractor(candidate_class, cspaces, matchers,
                                           self_relations=False, nested_relations=False,
                                           symmetric_relations=False)
```

An operator to extract Candidate objects from a Context.

### Parameters

- **candidate\_class** – The type of relation to extract, defined using `snorkel.models.candidate_subclass`
- **cspace** – one or list of `CandidateSpace` objects, one for each relation argument. Defines space of Contexts to consider
- **matchers** – one or list of `snorkel.matchers.Matcher` objects, one for each relation argument. Only tuples of Contexts for which each element is accepted by the corresponding Matcher will be returned as Candidates
- **self\_relations** – Boolean indicating whether to extract Candidates that relate the same context. Only applies to binary relations. Default is False.
- **nested\_relations** – Boolean indicating whether to extract Candidates that relate one Context with another that contains it. Only applies to binary relations. Default is False.
- **symmetric\_relations** – Boolean indicating whether to extract symmetric Candidates, i.e.,  $\text{rel}(A,B)$  and  $\text{rel}(B,A)$ , where A and B are Contexts. Only applies to binary relations. Default is False.

```
class snorkel.candidates.CandidateSpace
```

Defines the **space** of candidate objects Calling `_apply(x)_` given an object `_x_` returns a generator over candidates in `_x_`.

```
class snorkel.candidates.Ngrams(n_max=5, split_tokens=('-', '/'))
```

Defines the space of candidates as all n-grams ( $n \leq n\_max$ ) in a Sentence `_x_`, indexing by **character offset**.

```
class snorkel.candidates.PretaggedCandidateExtractor(candidate_class, entity_types, self_relations=False,
                                                    nested_relations=False, symmetric_relations=True,
                                                    entity_sep='~@~')
```

UDFRunner for PretaggedCandidateExtractorUDF

**class** snorkel.candidates.**PretaggedCandidateExtractorUDF** (*candidate\_class, entity\_types, self\_relations=False, nested\_relations=False, symmetric\_relations=False, entity\_sep='~@~', \*\*kwargs*)

An extractor for Sentences with entities pre-tagged, and stored in the `entity_types` and `entity_cids` fields.

**apply** (*context, clear, split, check\_for\_existing=True, \*\*kwargs*)  
Extract Candidates from a Context

**class** snorkel.matchers.**Concat** (*\*children, \*\*opts*)

Selects candidates which are the concatenation of adjacent matches from child operators NOTE: Currently slices on **word index** and considers concatenation along these divisions only

**class** snorkel.matchers.**DateMatcher** (*\*children, \*\*kwargs*)

Matches Spans that are dates, as identified by CoreNLP.

A convenience class for setting up a `RegexMatchEach` to match spans for which each token was tagged as a date.

**class** snorkel.matchers.**DictionaryMatch** (*\*children, \*\*opts*)

Selects candidate Ngrams that match against a given list `d`

**class** snorkel.matchers.**LambdaFunctionMatch** (*\*children, \*\*opts*)

Selects candidate Ngrams that match against a given list `d`

**class** snorkel.matchers.**LocationMatcher** (*\*children, \*\*kwargs*)

Matches Spans that are the names of locations, as identified by CoreNLP.

A convenience class for setting up a `RegexMatchEach` to match spans for which each token was tagged as a location.

**class** snorkel.matchers.**Matcher** (*\*children, \*\*opts*)

Applies a function `f: c -> {True,False}` to a generator of candidates, returning only candidates `_c_` s.t. `_f(c) == True_`, where `f` can be compositionally defined.

**apply** (*candidates*)

Apply the **Matcher** to a **generator** of candidates Optionally only takes the longest match (NOTE: assumes this is the *first* match)

**f** (*c*)

The recursively composed version of filter function `f` By default, returns logical **conjunction** of operator and single child operator

**class** snorkel.matchers.**MiscMatcher** (*\*children, \*\*kwargs*)

Matches Spans that are miscellaneous named entities, as identified by CoreNLP.

A convenience class for setting up a `RegexMatchEach` to match spans for which each token was tagged as miscellaneous.

**class** snorkel.matchers.**NgramMatcher** (*\*children, \*\*opts*)

Matcher base class for Ngram objects

**class** snorkel.matchers.**NumberMatcher** (*\*children, \*\*kwargs*)

Matches Spans that are numbers, as identified by CoreNLP.

A convenience class for setting up a `RegexMatchEach` to match spans for which each token was tagged as a number.

**class** snorkel.matchers.**OrganizationMatcher** (*\*children, \*\*kwargs*)

Matches Spans that are the names of organizations, as identified by CoreNLP.

A convenience class for setting up a `RegexMatchEach` to match spans for which each token was tagged as an organization.

**class** `snorkel.matchers.PersonMatcher` (\*children, \*\*kwargs)  
Matches Spans that are the names of people, as identified by CoreNLP.

A convenience class for setting up a `RegexMatchEach` to match spans for which each token was tagged as a person.

**class** `snorkel.matchers.RegexMatch` (\*children, \*\*opts)  
Base regex class- does not specify specific semantics of *what* is being matched yet

**class** `snorkel.matchers.RegexMatchEach` (\*children, \*\*opts)  
Matches regex pattern on **each token**

**class** `snorkel.matchers.RegexMatchSpan` (\*children, \*\*opts)  
Matches regex pattern on **full concatenated span**

**class** `snorkel.matchers.SlotFillMatch` (\*children, \*\*opts)  
Matches a slot fill pattern of matchers `_` at the character **level\_**

**class** `snorkel.matchers.Union` (\*children, \*\*opts)  
Takes the union of candidate sets returned by child operators

One of the core operations in Snorkel is `_annotating_` the candidates in various ways. We can think of generating features for the candidates as annotating them (creating a *Feature* object), and can also view supervision via labeling functions as annotating them (creating a *Label* object).

## Core Data Models

**class** `snorkel.models.annotation.AnnotationKeyMixin`

Mixin class for defining annotation key tables. An `AnnotationKey` is the unique name associated with a set of Annotations, corresponding e.g. to a single labeling or feature function. An `AnnotationKey` may have an associated weight (`Parameter`) associated with it.

**class** `snorkel.models.annotation.AnnotationMixin`

Mixin class for defining annotation tables. An annotation is a value associated with a `Candidate`. Examples include labels, features, and predictions. New types of annotations can be defined by creating an annotation class and corresponding annotation, for example:

```
from snorkel.models.annotation import AnnotationMixin
from snorkel.models.meta import SnorkelBase

class NewAnnotation(AnnotationMixin, SnorkelBase):
    value = Column(Float, nullable=False)

# The entire storage schema, including NewAnnotation, can now be initialized with
↳ the following import
import snorkel.models
```

The annotation class should include a `Column` attribute named `value`.

**class** `snorkel.models.annotation.Feature(**kwargs)`

An element of a representation of a `Candidate` in a feature space.

A Feature’s annotation key identifies the definition of the Feature, e.g., a function that implements it or the library name and feature name in an automatic featurization library.

**class** `snorkel.models.annotation.GoldLabel` (*\*\*kwargs*)  
 A separate class for labels from human annotators or other gold standards.

**class** `snorkel.models.annotation.Label` (*\*\*kwargs*)  
 A discrete label associated with a Candidate, indicating a target prediction value.  
 Labels are used to represent the output of labeling functions.

A Label’s annotation key identifies the labeling function that provided the Label.

**class** `snorkel.models.annotation.Prediction` (*\*\*kwargs*)  
 A probability associated with a Candidate, indicating the degree of belief that the Candidate is true.

A Prediction’s annotation key indicates which process or method produced the Prediction, e.g., which model with which ParameterSet.

**class** `snorkel.models.annotation.StableLabel` (*\*\*kwargs*)  
 A special secondary table for preserving labels created by *human annotators* (e.g. in the Viewer) in a stable format that does not cascade, and is independent of the Candidate ids.

## Core Objects for Annotations (Features, Labels)

**class** `snorkel.annotations.Annotator` (*annotation\_class, annotation\_key\_class, f\_gen*)  
 Abstract class for annotating candidates and persisting these annotations to DB

**apply\_existing** (*split=0, key\_group=0, cids\_query=None, \*\*kwargs*)  
 Alias for apply that emphasizes we are using an existing AnnotatorKey set.

**clear** (*session, split=0, key\_group=0, replace\_key\_set=True, cids\_query=None, \*\*kwargs*)  
 Deletes the Annotations for the Candidates in the given split. If *replace\_key\_set=True*, deletes *all* Annotations (of this Annotation sub-class) and also deletes all AnnotationKeys (of this sub-class)

**class** `snorkel.annotations.FeatureAnnotator` (*f=<function get\_span\_feats>*)  
 Apply feature generators to the candidates, generating Feature annotations

**class** `snorkel.annotations.LabelAnnotator` (*lfs=None, label\_generator=None*)  
 Apply labeling functions to the candidates, generating Label annotations

**Parameters** *lfs* – A *\_list\_* of labeling functions (LFs)

`snorkel.annotations.load_marginals` (*session, X=None, split=0, training=True*)  
 Load the marginal probs. for a given split of Candidates

`snorkel.annotations.load_matrix` (*matrix\_class, annotation\_key\_class, annotation\_class, session, split=0, cids\_query=None, key\_group=0, key\_names=None, zero\_one=False, load\_as\_array=False*)  
 Returns the annotations corresponding to a split of candidates with N members and an AnnotationKey group with M distinct keys as an N x M CSR sparse matrix.

`snorkel.annotations.save_marginals` (*session, X, marginals, training=True*)  
 Save marginal probabilities for a set of Candidates to db.

**Parameters**

- **X** – Either an M x N `csr_AnnotationMatrix`-class matrix, where M is number of candidates, N number of LFs/features; OR a list of arbitrary objects with candidate ids accessible via a `.id` attrib



- **marginals** – A dense  $M \times K$  matrix of marginal probabilities, where  $K$  is the cardinality of the candidates, OR a  $M$ -dim list/array if  $K=2$ .
- **training** – If True, these are training marginals / labels; else they are saved as end model predictions.

Note: The marginals for  $k=0$  are not stored, only for  $k = 1, \dots, K$



In the Snorkel pipeline, the user writes labeling functions (LFs) and then uses the generative model to unify and denoise their labels. Then, the marginal predictions of this model are used as probabilistic training labels for the discriminative model. Currently we provide bindings for TensorFlow models, and two basic models: Logistic regression and an LSTM. See tutorials for a more in-depth explanation.

## Base Classifier Class

**class** `snorkel.learning.classifier.Classifier` (*cardinality=2, name=None, seed=None*)  
Simple abstract base class for a probabilistic classifier.

**error\_analysis** (*session, X\_test, Y\_test, gold\_candidate\_set=None, b=0.5, set\_unlabeled\_as\_neg=True, display=True, scorer=<class 'snorkel.learning.utils.MentionScorer'>, \*\*kwargs*)

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

**load()**

**marginals** (*X*, *\*\*kwargs*)

**predictions** (*X*, *b=0.5*)

Return numpy array of elements in {-1,0,1} based on predicted marginal probabilities.

**representation = False**

**save()**

**save\_marginals** (*session*, *X*, *training=False*)

Save the predicted marginal probabilities for the Candidates *X*.

**score** (*X\_test*, *Y\_test*, *b=0.5*, *set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

**Parameters**

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*.

Note: Unlike in `self.error_analysis`, this method assumes `X_test` and `Y_test` are properly collated!

## Generative Model

```
class snorkel.learning.gen_learning.GenerativeModel (class_prior=False, lf_prior=False,
                                                    lf_propensity=False,
                                                    lf_class_propensity=False,
                                                    seed=271828, name=None)
```

A generative model for data programming for binary classification.

Supports dependencies among labeling functions.

**Parameters**

- **class\_prior** – whether to include class label prior factors
- **lf\_prior** – whether to include labeling function prior factors
- **lf\_propensity** – whether to include labeling function propensity factors
- **lf\_class\_propensity** – whether to include class-specific labeling function propensity factors
- **seed** – seed for initializing state of Numbskull variables

**dep\_names = ('dep\_similar', 'dep\_fixing', 'dep\_reinforcing', 'dep\_exclusive')**

```
error_analysis (session, X_test, Y_test, gold_candidate_set=None, b=0.5,
                set_unlabeled_as_neg=True, display=True, scorer=<class
                'snorkel.learning.utils.MentionScorer'>, **kwargs)
```

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

### **learned\_lf\_stats** ()

Provides a summary of what the model has learned about the labeling functions. For each labeling function, estimates of the following are provided:

Abstain Accuracy Coverage

[Following are only available for binary tasks] True Positive (TP) False Positive (FP) True Negative (TN) False Negative (FN)

For scoped categoricals, the information provided is for the maximum observed cardinality of any single data point.

**WARNING: This uses Gibbs sampling to estimate these values. This will** tend to mix poorly when there are many very accurate labeling functions. In this case, this function will assume that the classes are approximately balanced.

**load** (*model\_name=None, save\_dir='checkpoints', verbose=True*)

Load model.

**marginals** (*L, candidate\_ranges=None*)

Given an M x N label matrix, returns marginal probabilities for each candidate, depending on classification setting:

- Binary: Returns M-dim array representing the marginal probability** of each candidate being True
- Categorical (cardinality = K): Returns M x K dense matrix** representing the marginal probabilities of each candidate being each class.
- Scoped Categorical (cardinality = K, cardinality\_ranges not None):** Returns an M x K *sparse* matrix of marginals.

In the categorical setting, the K values (columns in the marginals matrix) correspond to indices of the Candidate values defined.

**optional\_names** = ('lf\_prior', 'lf\_propensity', 'lf\_class\_propensity')

**predictions** (*X, b=0.5*)

Return numpy array of elements in {-1,0,1} based on predicted marginal probabilities.

**representation** = False

**save** (*model\_name=None, save\_dir='checkpoints', verbose=True*)

Save current model.

**save\_marginals** (*session, X, training=False*)

Save the predicted marginal probabilities for the Candidates X.

**score** (*X\_test, Y\_test, b=0.5, set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

**Parameters**

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*.

Note: Unlike in `self.error_analysis`, this method assumes `X_test` and `Y_test` are properly collated!

**train** (*L, deps=(), LF\_acc\_prior\_weights=None, LF\_acc\_prior\_weight\_default=1, labels=None, label\_prior\_weight=5, init\_deps=0.0, init\_class\_prior=-1.0, epochs=30, step\_size=None, decay=1.0, reg\_param=0.1, reg\_type=2, verbose=False, truncation=10, burn\_in=5, cardinality=None, timer=None, candidate\_ranges=None, threads=1*)

Fits the parameters of the model to a data set. By default, learns a conditionally independent model. Additional unary dependencies can be set to be included in the constructor. Additional pairwise and higher-order dependencies can be included as an argument.

Results are stored as a member named `weights`, instance of `snorkel.learning.gen_learning.GenerativeModelWeights`.

**Parameters**

- **L** – M x N `csr_ANNOTATIONMATRIX`-type label matrix, where there are M candidates labeled by N labeling functions (LFs)
- **deps** – collection of dependencies to include in the model, each element is a tuple of the form (LF 1 index, LF 2 index, dependency type), see `snorkel.learning.constants`
- **LF\_acc\_prior\_weights** – An N-element list of prior weights for the LF accuracies (log scale)
- **LF\_acc\_prior\_weight\_default** – Default prior for the weight of each LF accuracy; if `LF_acc_prior_weights` is unset, each LF will have this accuracy prior weight (log scale)
- **labels** – Optional ground truth labels
- **label\_prior\_weight** – The prior probability that the ground truth labels (if provided) are correct (log scale)
- **init\_deps** – initial weight for additional dependencies, except class prior (log scale)
- **init\_class\_prior** – initial class prior (in log scale), note only used if `class_prior=True` in constructor
- **epochs** – number of training epochs
- **step\_size** – gradient step size, default is  $1 / L.shape[0]$
- **decay** – multiplicative decay of step size,  $step\_size_{(t+1)} = step\_size_{(t)} * decay$
- **reg\_param** – regularization strength

- **reg\_type** – 1 = L1 regularization, 2 = L2 regularization
- **verbose** – whether to write debugging info to stdout
- **truncation** – number of iterations between truncation step for L1 regularization
- **burn\_in** – number of burn-in samples to take before beginning learning
- **cardinality** – number of possible classes; by default is inferred from the label matrix L
- **timer** – stopwatch for profiling, must implement start() and end()
- **candidate\_ranges** – Optionally, a list of M sets of integer values, representing the possible categorical values that each of the M candidates can take. If a label is outside of this range throws an error. If None, then each candidate can take any value from 0 to cardinality.
- **threads** – the number of threads to use for sampling. Default is 1.

**class** snorkel.learning.gen\_learning.**GenerativeModelWeights** (*n*)

**is\_sign\_sparsistent** (*other*, *threshold=0.1*)

## Discriminative Models

**class** snorkel.learning.disc\_learning.**TFNoiseAwareModel** (*n\_threads=None*, *\*\*kwargs*)

Generic NoiseAwareModel class for TensorFlow models. Note that the actual network is built when train is called (to allow for model architectures which depend on the training data, e.g. vocab size).

**Parameters** **n\_threads** – Parallelism to use; single-threaded if None

**error\_analysis** (*session*, *X\_test*, *Y\_test*, *gold\_candidate\_set=None*, *b=0.5*,  
*set\_unlabeled\_as\_neg=True*, *display=True*, *scorer=<class*  
*'snorkel.learning.utils.MentionScorer'>*, *\*\*kwargs*)

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

**load** (*model\_name=None*, *save\_dir='checkpoints'*, *verbose=True*)

Load model from file and rebuild in new graph / session.

**marginals** (*X*, *\*\*kwargs*)

**predictions** (*X*, *b=0.5*)

Return numpy array of elements in  $\{-1,0,1\}$  based on predicted marginal probabilities.

**representation = False**

**save** (*model\_name=None*, *save\_dir='checkpoints'*, *verbose=True*, *global\_step=0*)

Save current model.

**save\_marginals** (*session*, *X*, *training=False*)

Save the predicted marginal probabilities for the Candidates *X*.

**score** (*X\_test*, *Y\_test*, *b=0.5*, *set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels  $\rightarrow -1$ , *binary setting*.

Note: Unlike in `self.error_analysis`, this method assumes `X_test` and `Y_test` are properly collated!

**train** (*X\_train*, *Y\_train*, *n\_epochs=25*, *lr=0.01*, *batch\_size=256*, *rebalance=False*, *X\_dev=None*, *Y\_dev=None*, *print\_freq=5*, *dev\_ckpt=True*, *dev\_ckpt\_delay=0.75*, *save\_dir='checkpoints'*, *\*\*kwargs*)

Generic training procedure for TF model

#### Parameters

- **X\_train** – The training Candidates. If `self.representation` is `True`, then this is a list of Candidate objects; else is a `csr_ANNOTATIONMATRIX` with rows corresponding to training candidates and columns corresponding to features.
- **Y\_train** – Array of marginal probabilities for each Candidate
- **n\_epochs** – Number of training epochs
- **lr** – Learning rate
- **batch\_size** – Batch size for SGD
- **rebalance** – Bool or fraction of positive examples for training - if `True`, defaults to standard 0.5 class balance - if `False`, no class balancing
- **X\_dev** – Candidates for evaluation, same format as `X_train`
- **Y\_dev** – Labels for evaluation, same format as `Y_train`
- **print\_freq** – number of epochs at which to print status, and if present, evaluate the dev set (`X_dev`, `Y_dev`).
- **dev\_ckpt** – If `True`, save a checkpoint whenever highest score on (`X_dev`, `Y_dev`) reached. Note: currently only evaluates at every `@print_freq` epochs.
- **dev\_ckpt\_delay** – Start dev checkpointing after this portion of `n_epochs`.
- **save\_dir** – Save dir path for checkpointing.



- **kwargs** – All hyperparameters that change how the graph is built must be passed through here to be saved and reloaded to save / reload model. *NOTE: If a parameter needed to build the network and/or is needed at test time is not included here, the model will not be able to be reloaded!*

**class** snorkel.learning.disc\_models.logistic\_regression.**LogisticRegression** (*n\_threads=None, \*\*kwargs*)

**error\_analysis** (*session, X\_test, Y\_test, gold\_candidate\_set=None, b=0.5, set\_unlabeled\_as\_neg=True, display=True, scorer=<class 'snorkel.learning.utils.MentionScorer'>, \*\*kwargs*)

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

**get\_weights** ()

Get model weights and bias

**load** (*model\_name=None, save\_dir='checkpoints', verbose=True*)

Load model from file and rebuild in new graph / session.

**marginals** (*X\_test*)

**predictions** (*X, b=0.5*)

Return numpy array of elements in {-1,0,1} based on predicted marginal probabilities.

**representation = False**

**save** (*model\_name=None, save\_dir='checkpoints', verbose=True, global\_step=0*)

Save current model.

**save\_marginals** (*session, X, training=False*)

Save the predicted marginal probabilities for the Candidates X.

**score** (*X\_test, Y\_test, b=0.5, set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix

- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*.

Note: Unlike in `self.error_analysis`, this method assumes `X_test` and `Y_test` are properly collated!

```
train(X_train, Y_train, n_epochs=25, lr=0.01, batch_size=256, rebalance=False, X_dev=None,
      Y_dev=None, print_freq=5, dev_ckpt=True, dev_ckpt_delay=0.75, save_dir='checkpoints',
      **kwargs)
```

Generic training procedure for TF model

#### Parameters

- **X\_train** – The training Candidates. If `self.representation` is `True`, then this is a list of Candidate objects; else is a `csr_AnnotationMatrix` with rows corresponding to training candidates and columns corresponding to features.
- **Y\_train** – Array of marginal probabilities for each Candidate
- **n\_epochs** – Number of training epochs
- **lr** – Learning rate
- **batch\_size** – Batch size for SGD
- **rebalance** – Bool or fraction of positive examples for training - if `True`, defaults to standard 0.5 class balance - if `False`, no class balancing
- **X\_dev** – Candidates for evaluation, same format as `X_train`
- **Y\_dev** – Labels for evaluation, same format as `Y_train`
- **print\_freq** – number of epochs at which to print status, and if present, evaluate the dev set (`X_dev`, `Y_dev`).
- **dev\_ckpt** – If `True`, save a checkpoint whenever highest score on (`X_dev`, `Y_dev`) reached. Note: currently only evaluates at every `@print_freq` epochs.
- **dev\_ckpt\_delay** – Start dev checkpointing after this portion of `n_epochs`.
- **save\_dir** – Save dir path for checkpointing.
- **kwargs** – All hyperparameters that change how the graph is built must be passed through here to be saved and reloaded to save / reload model. *NOTE: If a parameter needed to build the network and/or is needed at test time is not included here, the model will not be able to be reloaded!*

```
class snorkel.learning.disc_models.logistic_regression.SparseLogisticRegression(n_threads=None,
                                                                              **kwargs)
```

```
error_analysis(session, X_test, Y_test, gold_candidate_set=None, b=0.5,
               set_unlabeled_as_neg=True, display=True, scorer=<class
               'snorkel.learning.utils.MentionScorer'>, **kwargs)
```

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix

- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

**get\_weights** ()

Get model weights and bias

**load** (*model\_name=None, save\_dir='checkpoints', verbose=True*)

Load model from file and rebuild in new graph / session.

**marginals** (*X\_test*)

**predictions** (*X, b=0.5*)

Return numpy array of elements in {-1,0,1} based on predicted marginal probabilities.

**representation = False**

**save** (*model\_name=None, save\_dir='checkpoints', verbose=True, global\_step=0*)

Save current model.

**save\_marginals** (*session, X, training=False*)

Save the predicted marginal probabilities for the Candidates X.

**score** (*X\_test, Y\_test, b=0.5, set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

**Parameters**

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*.

Note: Unlike in `self.error_analysis`, this method assumes `X_test` and `Y_test` are properly collated!

**train** (*X\_train, Y\_train, n\_epochs=25, lr=0.01, batch\_size=256, rebalance=False, X\_dev=None, Y\_dev=None, print\_freq=5, dev\_ckpt=True, dev\_ckpt\_delay=0.75, save\_dir='checkpoints', \*\*kwargs*)

Generic training procedure for TF model

**Parameters**

- **X\_train** – The training Candidates. If `self.representation` is `True`, then this is a list of Candidate objects; else is a `csr_ANNOTATIONMATRIX` with rows corresponding to training candidates and columns corresponding to features.
- **Y\_train** – Array of marginal probabilities for each Candidate
- **n\_epochs** – Number of training epochs
- **lr** – Learning rate

- **batch\_size** – Batch size for SGD
- **rebalance** – Bool or fraction of positive examples for training - if True, defaults to standard 0.5 class balance - if False, no class balancing
- **X\_dev** – Candidates for evaluation, same format as X\_train
- **Y\_dev** – Labels for evaluation, same format as Y\_train
- **print\_freq** – number of epochs at which to print status, and if present, evaluate the dev set (X\_dev, Y\_dev).
- **dev\_ckpt** – If True, save a checkpoint whenever highest score on (X\_dev, Y\_dev) reached. Note: currently only evaluates at every @print\_freq epochs.
- **dev\_ckpt\_delay** – Start dev checkpointing after this portion of n\_epochs.
- **save\_dir** – Save dir path for checkpointing.
- **kwargs** – All hyperparameters that change how the graph is built must be passed through here to be saved and reloaded to save / reload model. *NOTE: If a parameter needed to build the network and/or is needed at test time is not included here, the model will not be able to be reloaded!*

**class** snorkel.learning.disc\_models.rnn.rnn\_base.**RNNBase** (*n\_threads=None, \*\*kwargs*)

**error\_analysis** (*session, X\_test, Y\_test, gold\_candidate\_set=None, b=0.5, set\_unlabeled\_as\_neg=True, display=True, scorer=<class 'snorkel.learning.utils.MentionScorer'>, \*\*kwargs*)

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

**load** (*model\_name=None, save\_dir='checkpoints', verbose=True*)  
Load model from file and rebuild in new graph / session.

**marginals** (*test\_candidates*)

Get likelihood of tagged sequences represented by test\_candidates @test\_candidates: list of lists representing test sentence

**predictions** (*X, b=0.5*)

Return numpy array of elements in {-1,0,1} based on predicted marginal probabilities.

**representation = True**

**save** (*model\_name=None, save\_dir='checkpoints', verbose=True, global\_step=0*)  
Save current model.

**save\_marginals** (*session, X, training=False*)  
Save the predicted marginal probabilities for the Candidates X.

**score** (*X\_test, Y\_test, b=0.5, set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

**Parameters**

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*.

Note: Unlike in `self.error_analysis`, this method assumes `X_test` and `Y_test` are properly collated!

**train** (*X\_train, Y\_train, X\_dev=None, max\_sentence\_length=None, \*\*kwargs*)  
Perform preprocessing of data, construct dataset-specific model, then train.

`snorkel.learning.disc_models.rnn.re_rnn.mark` (*l, h, idx*)  
Produce markers based on argument positions

**Parameters**

- **l** – sentence position of first word in argument
- **h** – sentence position of last word in argument
- **idx** – argument index (1 or 2)

`snorkel.learning.disc_models.rnn.re_rnn.mark_sentence` (*s, args*)  
Insert markers around relation arguments in word sequence

**Parameters**

- **s** – list of tokens in sentence
- **args** – list of triples (l, h, idx) as per `@_mark(...)` corresponding to relation arguments

**Example: Then Barack married Michelle.** -> Then ~-[1 Barack 1]~- married ~-[2 Michelle 2]~-.

**class** `snorkel.learning.disc_models.rnn.re_rnn.reRNN` (*n\_threads=None, \*\*kwargs*)  
reRNN for relation extraction

**error\_analysis** (*session, X\_test, Y\_test, gold\_candidate\_set=None, b=0.5, set\_unlabeled\_as\_neg=True, display=True, scorer=<class 'snorkel.learning.utils.MentionScorer'>, \*\*kwargs*)

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

**Parameters**

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

**load** (*model\_name=None, save\_dir='checkpoints', verbose=True*)  
 Load model from file and rebuild in new graph / session.

**marginals** (*test\_candidates*)  
 Get likelihood of tagged sequences represented by test\_candidates @test\_candidates: list of lists representing test sentence

**predictions** (*X, b=0.5*)  
 Return numpy array of elements in {-1,0,1} based on predicted marginal probabilities.

**representation = True**

**save** (*model\_name=None, save\_dir='checkpoints', verbose=True, global\_step=0*)  
 Save current model.

**save\_marginals** (*session, X, training=False*)  
 Save the predicted marginal probabilities for the Candidates X.

**score** (*X\_test, Y\_test, b=0.5, set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

**Parameters**

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*.

Note: Unlike in self.error\_analysis, this method assumes X\_test and Y\_test are properly collated!

**train** (*X\_train, Y\_train, X\_dev=None, max\_sentence\_length=None, \*\*kwargs*)  
 Perform preprocessing of data, construct dataset-specific model, then train.

**class** snorkel.learning.disc\_models.rnn.tag\_rnn.**TagRNN** (*n\_threads=None, \*\*kwargs*)  
 TagRNN for sequence tagging

**CLOSE** = '~~]]~~'

**OPEN** = '~~[[~~'

**error\_analysis** (*session*, *X\_test*, *Y\_test*, *gold\_candidate\_set=None*, *b=0.5*,  
*set\_unlabeled\_as\_neg=True*, *display=True*, *scorer=<class  
'snorkel.learning.utils.MentionScorer'>*, *\*\*kwargs*)

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

**load** (*model\_name=None*, *save\_dir='checkpoints'*, *verbose=True*)  
Load model from file and rebuild in new graph / session.

**marginals** (*test\_candidates*)

Get likelihood of tagged sequences represented by test\_candidates @test\_candidates: list of lists representing test sentence

**predictions** (*X*, *b=0.5*)

Return numpy array of elements in {-1,0,1} based on predicted marginal probabilities.

**representation = True**

**save** (*model\_name=None*, *save\_dir='checkpoints'*, *verbose=True*, *global\_step=0*)  
Save current model.

**save\_marginals** (*session*, *X*, *training=False*)

Save the predicted marginal probabilities for the Candidates X.

**score** (*X\_test*, *Y\_test*, *b=0.5*, *set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*.

Note: Unlike in self.error\_analysis, this method assumes X\_test and Y\_test are properly collated!

**train** (*X\_train*, *Y\_train*, *X\_dev=None*, *max\_sentence\_length=None*, *\*\*kwargs*)  
Perform preprocessing of data, construct dataset-specific model, then train.

`snorkel.learning.disc_models.rnn.tag_rnn.tag` (*seq, labels*)

**class** `snorkel.learning.disc_models.rnn.text_rnn.TextRNN` (*n\_threads=None, \*\*kwargs*)  
 TextRNN for strings of text.

**error\_analysis** (*session, X\_test, Y\_test, gold\_candidate\_set=None, b=0.5, set\_unlabeled\_as\_neg=True, display=True, scorer=<class 'snorkel.learning.utils.MentionScorer'>, \*\*kwargs*)

Prints full score analysis using the Scorer class, and then returns the a tuple of sets conatining the test candidates bucketed for error analysis, i.e.:

- For binary: TP, FP, TN, FN
- For categorical: correct, incorrect

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **gold\_candidate\_set** – Full set of TPs in the test set
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*
- **display** – Print score report
- **scorer** – The Scorer sub-class to use

**load** (*model\_name=None, save\_dir='checkpoints', verbose=True*)  
 Load model from file and rebuild in new graph / session.

**marginals** (*test\_candidates*)

Get likelihood of tagged sequences represented by test\_candidates @test\_candidates: list of lists representing test sentence

**predictions** (*X, b=0.5*)

Return numpy array of elements in {-1,0,1} based on predicted marginal probabilities.

**representation = True**

**save** (*model\_name=None, save\_dir='checkpoints', verbose=True, global\_step=0*)  
 Save current model.

**save\_marginals** (*session, X, training=False*)

Save the predicted marginal probabilities for the Candidates X.

**score** (*X\_test, Y\_test, b=0.5, set\_unlabeled\_as\_neg=True*)

**Returns the summary scores:**

- For binary: precision, recall, F1 score
- For categorical: accuracy

#### Parameters

- **X\_test** – The input test candidates, as a list or annotation matrix
- **Y\_test** – The input test labels, as a list or annotation matrix
- **b** – Decision boundary *for binary setting only*
- **set\_unlabeled\_as\_neg** – Whether to map 0 labels -> -1, *binary setting*.



Note: Unlike in `self.error_analysis`, this method assumes `X_test` and `Y_test` are properly collated!

**train** (`X_train`, `Y_train`, `X_dev=None`, `max_sentence_length=None`, *\*\*kwargs*)  
 Perform preprocessing of data, construct dataset-specific model, then train.

## Learning Utilities

**class** `snorkel.learning.utils.GridSearch` (`model_class`, `parameters`, `X_train`, `Y_train=None`, *\*\*model\_class\_params*)  
 Runs hyperparameter grid search over a model object with `train` and `score` methods, training data (`X`), and `training_marginals`. Selects based on maximizing F1 score on a supplied validation set. Specify search space with Hyperparameter arguments.

**fit** (`X_valid`, `Y_valid`, `b=0.5`, `set_unlabeled_as_neg=True`, `validation_kwargs={}`, `n_threads=1`, `save_dir='checkpoints'`, *\*\*model\_hyperparams*)

**search\_space** ()

**class** `snorkel.learning.utils.Hyperparameter` (`name`)  
 Base class for a grid search parameter

**draw\_values** (`n`)

**get\_all\_values** ()

`snorkel.learning.utils.LF_accuracies` (`L`, `labels`)  
 Given an `N x M` matrix where `L_{i,j}` is the label given by the `j`th LF to the `i`th candidate, and `labels` `{-1,1}`. Return the accuracy of each LF w.r.t. these labels.

`snorkel.learning.utils.LF_conflicts` (`L`)  
 Given an `N x M` matrix where `L_{i,j}` is the label given by the `j`th LF to the `i`th candidate: Return the **fraction of candidates that each LF conflicts with other LFs on**.

`snorkel.learning.utils.LF_coverage` (`L`)  
 Given an `N x M` matrix where `L_{i,j}` is the label given by the `j`th LF to the `i`th candidate: Return the **fraction of candidates that each LF labels**.

`snorkel.learning.utils.LF_overlaps` (`L`)  
 Given an `N x M` matrix where `L_{i,j}` is the label given by the `j`th LF to the `i`th candidate: Return the **fraction of candidates that each LF overlaps with other LFs on**.

**class** `snorkel.learning.utils.LabelBalancer` (`y`)

**get\_train\_idxs** (`rebalance=False`, `split=0.5`)  
 Get training indices based on `@y`. `@rebalance`: bool or fraction of positive examples desired.  
 If True, default fraction is 0.5. If False no balancing.  
`@split`: Split point for positive and negative classes.

**class** `snorkel.learning.utils.ListParameter` (`name`, `parameter_list`)  
 List of parameter values for searching

**draw\_values** (`n`)

**get\_all\_values** ()

**class** `snorkel.learning.utils.MentionScorer` (`test_candidates`, `test_labels`, `gold_candidate_set=None`)  
 Scorer for mention level assessment

**score** (*test\_marginals*, *\*\*kwargs*)

**summary\_score** (*test\_marginals*, *\*\*kwargs*)

Return the F1 score (for binary) or accuracy (for categorical). Also return the label as second argument.

**class** `snorkel.learning.utils.ModelTester` (*model\_class*, *model\_class\_params*, *params\_queue*, *scores\_queue*, *X\_train*, *X\_valid*, *Y\_valid*, *Y\_train=None*, *b=0.5*, *set\_unlabeled\_as\_neg=True*, *save\_dir='checkpoints'*)

**authkey**

**daemon**

Return whether process is a daemon

**exitcode**

Return exit code of process or *None* if it has yet to stop

**ident**

Return identifier (PID) of process or *None* if it has yet to start

**is\_alive** ()

Return whether process is alive

**join** (*timeout=None*)

Wait until child process terminates

**name**

**pid**

Return identifier (PID) of process or *None* if it has yet to start

**run** ()

**start** ()

Start child process

**terminate** ()

Terminate process; sends SIGTERM signal or uses `TerminateProcess()`

**class** `snorkel.learning.utils.RandomSearch` (*model\_class*, *parameters*, *X\_train*, *Y\_train=None*, *n=10*, *\*\*model\_class\_params*)

**fit** (*X\_valid*, *Y\_valid*, *b=0.5*, *set\_unlabeled\_as\_neg=True*, *validation\_kwargs={}*, *n\_threads=1*, *save\_dir='checkpoints'*, *\*\*model\_hyperparams*)

**search\_space** ()

**class** `snorkel.learning.utils.RangeParameter` (*name*, *v1*, *v2*, *step=1*, *log\_base=None*)

Range of parameter values for searching. *min\_value* and *max\_value* are the ends of the search range. If *log\_base* is specified, scale the search range in the log base. *step* is range step size or exponent step size.

**draw\_values** (*n*)

**get\_all\_values** ()

**class** `snorkel.learning.utils.Scorer` (*test\_candidates*, *test\_labels*, *gold\_candidate\_set=None*)

Abstract type for scorers

**score** (*test\_marginals*, *\*\*kwargs*)

**summary\_score** (*test\_marginals*, *\*\*kwargs*)

Return the F1 score (for binary) or accuracy (for categorical).

`snorkel.learning.utils.binary_scores_from_counts` (*ntp, nfp, ntn, nfn*)

Precision, recall, and F1 scores from counts of TP, FP, TN, FN. Example usage:

```
p, r, f1 = binary_scores_from_counts(*map(len, error_sets))
```

`snorkel.learning.utils.calibration_plots` (*train\_marginals,* *test\_marginals,*  
*gold\_labels=None*)

Show classification accuracy and probability histogram plots

`snorkel.learning.utils.candidate_conflict` (*L*)

Given an  $N \times M$  matrix where  $L_{\{i,j\}}$  is the label given by the  $j$ th LF to the  $i$ th candidate: Return the **fraction of candidates which have > 1 (non-zero) labels \_which are not equal\_**.

`snorkel.learning.utils.candidate_coverage` (*L*)

Given an  $N \times M$  matrix where  $L_{\{i,j\}}$  is the label given by the  $j$ th LF to the  $i$ th candidate: Return the **fraction of candidates which have > 0 (non-zero) labels**.

`snorkel.learning.utils.candidate_overlap` (*L*)

Given an  $N \times M$  matrix where  $L_{\{i,j\}}$  is the label given by the  $j$ th LF to the  $i$ th candidate: Return the **fraction of candidates which have > 1 (non-zero) labels**.

`snorkel.learning.utils.plot_accuracy` (*probs, ground\_truth*)

`snorkel.learning.utils.plot_prediction_probability` (*probs*)

`snorkel.learning.utils.print_scores` (*ntp, nfp, ntn, nfn, title='Scores'*)

`snorkel.learning.utils.reshape_marginals` (*marginals*)

Returns correctly shaped marginals as np array

`snorkel.learning.utils.sparse_abs` (*X*)

Element-wise absolute value of sparse matrix- avoids casting to dense matrix!

`snorkel.learning.utils.training_set_summary_stats` (*L,* *return\_vals=True,* *ver-*  
*bose=False*)

Given an  $N \times M$  matrix where  $L_{\{i,j\}}$  is the label given by the  $j$ th LF to the  $i$ th candidate: Return simple summary statistics



---

## Etc: Viewing and Annotating Data, Writing LFs

---

### Using the Viewer to Inspect and Annotate Data

```

class snorkel.viewer.SentenceNgramViewer(candidates, session, gold=[], n_per_page=3,
                                         height=225, annotator_name=None)
    Viewer for Sentence objects and candidate Spans within them

class snorkel.viewer.Viewer(candidates, session, gold=[], n_per_page=3, height=225, annota-
                           tor_name=None)
    Generic object for viewing and labeling Candidate objects in their rendered Contexts.

    handle_label_event (_, content, buffers)
        Handles label event by persisting new label

    render ()
        Renders viewer pane

```

### Helpers for Writing Labeling Functions

```

snorkel.lf_helpers.contains_token(c, tok, attrib='words', case_sensitive=False)
    Checks if any of the constituent Spans contain a token :param attrib: The token attribute type (e.g. words,
    lemmas, poses)

snorkel.lf_helpers.get_between_tokens(c, attrib='words', n_max=1, case_sensitive=False)
    TODO: write doc_string

snorkel.lf_helpers.get_doc_candidate_spans(c)
    Get the Spans in the same document as Candidate c, where these Spans are arguments of Candidates.

snorkel.lf_helpers.get_left_tokens(c, window=3, attrib='words', n_max=1,
                                   case_sensitive=False)
    Return the tokens within a window to the _left_ of the Candidate. For higher-arity Candidates, defaults to the
    _first_argument_. :param window: The number of tokens to the left of the first argument to

    return

```

**Parameters** `attrib` – The token attribute type (e.g. words, lemmas, poses)

`snorkel.lf_helpers.get_matches(lf, candidate_set, match_values=[1, -1])`

A simple helper function to see how many matches (non-zero by default) an LF gets. Returns the matched set, which can then be directly put into the Viewer.

`snorkel.lf_helpers.get_right_tokens(c, window=3, attrib='words', n_max=1, case_sensitive=False)`

Return the tokens within a window to the `_right_` of the Candidate. For higher-arity Candidates, defaults to the `_last_` argument. :param window: The number of tokens to the right of the last argument to

return

**Parameters** `attrib` – The token attribute type (e.g. words, lemmas, poses)

`snorkel.lf_helpers.get_sent_candidate_spans(c)`

Get the Spans in the same Sentence as Candidate `c`, where these Spans are arguments of Candidates.

`snorkel.lf_helpers.get_tagged_text(c)`

Returns the text of `c`'s parent context with `c`'s unary spans replaced with tags `{{A}}`, `{{B}}`, etc. A convenience method for writing LFs based on e.g. regexes.

`snorkel.lf_helpers.get_text_between(c)`

Returns the text between the two unary Spans of a binary-Span Candidate, where both are in the same Sentence.

`snorkel.lf_helpers.get_text_splits(c)`

Given a k-arity Candidate defined over k Spans, return the chunked parent context (e.g. Sentence) split around the k constituent Spans.

NOTE: Currently assumes that these Spans are in the same Context

`snorkel.lf_helpers.is_inverted(c)`

Returns True if the ordering of the candidates in the sentence is inverted.

`snorkel.lf_helpers.test_LF(session, lf, split, annotator_name)`

Gets the accuracy of a single LF on a split of the candidates, w.r.t. annotator labels, and also returns the error buckets of the candidates.

## Helpers for Loading External Annotations

## S

`snorkel.annotations`, 12  
`snorkel.candidates`, 8  
`snorkel.learning.classifier`, 15  
`snorkel.learning.disc_learning`, 19  
`snorkel.learning.disc_models.logistic_regression`,  
21  
`snorkel.learning.disc_models.rnn.re_rnn`,  
25  
`snorkel.learning.disc_models.rnn.rnn_base`,  
24  
`snorkel.learning.disc_models.rnn.tag_rnn`,  
26  
`snorkel.learning.disc_models.rnn.text_rnn`,  
28  
`snorkel.learning.gen_learning`, 16  
`snorkel.learning.utils`, 29  
`snorkel.lf_helpers`, 33  
`snorkel.matchers`, 9  
`snorkel.models.annotation`, 11  
`snorkel.models.candidate`, 7  
`snorkel.models.context`, 3  
`snorkel.parser.corpus_parser`, 4  
`snorkel.parser.doc_preprocessors`, 4  
`snorkel.parser.parser`, 5  
`snorkel.viewer`, 33





## A

AnnotationKeyMixin (class in snorkel.models.annotation), 11  
 AnnotationMixin (class in snorkel.models.annotation), 11  
 Annotator (class in snorkel.annotations), 12  
 apply() (snorkel.candidates.PretaggedCandidateExtractorUDPipe method), 9  
 apply() (snorkel.matchers.Matcher method), 9  
 apply\_existing() (snorkel.annotations.Annotator method), 12  
 authkey (snorkel.learning.utils.ModelTester attribute), 30

## B

binary\_scores\_from\_counts() (in module snorkel.learning.utils), 30

## C

calibration\_plots() (in module snorkel.learning.utils), 31  
 Candidate (class in snorkel.models.candidate), 7  
 candidate\_conflict() (in module snorkel.learning.utils), 31  
 candidate\_coverage() (in module snorkel.learning.utils), 31  
 candidate\_overlap() (in module snorkel.learning.utils), 31  
 candidate\_subclass() (in module snorkel.models.candidate), 7  
 CandidateExtractor (class in snorkel.candidates), 8  
 CandidateSpace (class in snorkel.candidates), 8  
 char\_to\_word\_index() (snorkel.models.context.TemporarySpan method), 3  
 Classifier (class in snorkel.learning.classifier), 15  
 clear() (snorkel.annotations.Annotator method), 12  
 CLOSE (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN attribute), 26  
 Concat (class in snorkel.matchers), 9  
 construct\_stable\_id() (in module snorkel.models.context), 4  
 contains\_token() (in module snorkel.lf\_helpers), 33  
 Context (class in snorkel.models.context), 3

CSVPathsPreprocessor (class in snorkel.parser.doc\_preprocessors), 4

## D

daemon (snorkel.learning.utils.ModelTester attribute), 30  
 DateMatcher (class in snorkel.matchers), 9  
 dep\_names (snorkel.learning.gen\_learning.GenerativeModel attribute), 16  
 DictionaryMatch (class in snorkel.matchers), 9  
 DocPreprocessor (class in snorkel.parser.doc\_preprocessors), 4  
 Document (class in snorkel.models.context), 3  
 draw\_values() (snorkel.learning.utils.Hyperparameter method), 29  
 draw\_values() (snorkel.learning.utils.ListParameter method), 29  
 draw\_values() (snorkel.learning.utils.RangeParameter method), 30

## E

error\_analysis() (snorkel.learning.classifier.Classifier method), 15  
 error\_analysis() (snorkel.learning.disc\_learning.TFNoiseAwareModel method), 19  
 error\_analysis() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression method), 21  
 error\_analysis() (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression method), 22  
 error\_analysis() (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN method), 25  
 error\_analysis() (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase method), 24  
 error\_analysis() (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN method), 26  
 error\_analysis() (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN method), 28  
 error\_analysis() (snorkel.learning.gen\_learning.GenerativeModel method), 16  
 exitcode (snorkel.learning.utils.ModelTester attribute), 30

## F

- f() (snorkel.matchers.Matcher method), 9
- Feature (class in snorkel.models.annotation), 11
- FeatureAnnotator (class in snorkel.annotations), 12
- fit() (snorkel.learning.utils.GridSearch method), 29
- fit() (snorkel.learning.utils.RandomSearch method), 30

## G

- generate() (snorkel.parser.doc\_preprocessors.DocPreprocessor method), 4
- GenerativeModel (class in snorkel.learning.gen\_learning), 16
- GenerativeModelWeights (class in snorkel.learning.gen\_learning), 19
- get\_all\_values() (snorkel.learning.utils.Hyperparameter method), 29
- get\_all\_values() (snorkel.learning.utils.ListParameter method), 29
- get\_all\_values() (snorkel.learning.utils.RangeParameter method), 30
- get\_attrib\_span() (snorkel.models.context.TemporarySpan method), 3
- get\_attrib\_tokens() (snorkel.models.context.TemporarySpan method), 4
- get\_between\_tokens() (in module snorkel.lf\_helpers), 33
- get\_cids() (snorkel.models.candidate.Candidate method), 7
- get\_contexts() (snorkel.models.candidate.Candidate method), 7
- get\_doc\_candidate\_spans() (in module snorkel.lf\_helpers), 33
- get\_left\_tokens() (in module snorkel.lf\_helpers), 33
- get\_matches() (in module snorkel.lf\_helpers), 34
- get\_right\_tokens() (in module snorkel.lf\_helpers), 34
- get\_sent\_candidate\_spans() (in module snorkel.lf\_helpers), 34
- get\_tagged\_text() (in module snorkel.lf\_helpers), 34
- get\_text\_between() (in module snorkel.lf\_helpers), 34
- get\_text\_splits() (in module snorkel.lf\_helpers), 34
- get\_train\_idxs() (snorkel.learning.utils.LabelBalancer method), 29
- get\_weights() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression class in snorkel.matchers), 9
- get\_weights() (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression class in snorkel.matchers), 9
- GoldLabel (class in snorkel.models.annotation), 12
- GridSearch (class in snorkel.learning.utils), 29

## H

- handle\_label\_event() (snorkel.viewer.Viewer method), 33
- HTMLDocPreprocessor (class in snorkel.parser.doc\_preprocessors), 4
- Hyperparameter (class in snorkel.learning.utils), 29

## I

- ident (snorkel.learning.utils.ModelTester attribute), 30
- is\_alive() (snorkel.learning.utils.ModelTester method), 30
- is\_inverted() (in module snorkel.lf\_helpers), 34
- is\_sign\_sparsistent() (snorkel.learning.gen\_learning.GenerativeModelWeights method), 19

## J

- join() (snorkel.learning.utils.ModelTester method), 30

## L

- Label (class in snorkel.models.annotation), 12
- LabelAnnotator (class in snorkel.annotations), 12
- LabelBalancer (class in snorkel.learning.utils), 29
- LambdaFunctionMatch (class in snorkel.matchers), 9
- learned\_lf\_stats() (snorkel.learning.gen\_learning.GenerativeModel method), 17
- LF\_accuracies() (in module snorkel.learning.utils), 29
- LF\_conflicts() (in module snorkel.learning.utils), 29
- LF\_coverage() (in module snorkel.learning.utils), 29
- LF\_overlaps() (in module snorkel.learning.utils), 29
- ListParameter (class in snorkel.learning.utils), 29
- load() (snorkel.learning.classifier.Classifier method), 15
- load() (snorkel.learning.disc\_learning.TFNoiseAwareModel method), 19
- load() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression method), 21
- load() (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression method), 23
- load() (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN method), 26
- load() (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase method), 24
- load() (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN method), 27
- load() (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN method), 28
- load() (snorkel.learning.gen\_learning.GenerativeModel method), 17
- load\_marginals() (in module snorkel.annotations), 12
- load\_matrix() (in module snorkel.annotations), 12
- LocationMatcher (class in snorkel.matchers), 9
- LogisticRegression (class in snorkel.learning.disc\_models.logistic\_regression), 21

## M

- Marginal (class in snorkel.models.candidate), 7
- marginals() (snorkel.learning.classifier.Classifier method), 16
- marginals() (snorkel.learning.disc\_learning.TFNoiseAwareModel method), 19
- marginals() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression method), 21

marginals() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression method), 23  
 marginals() (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN method), 26  
 marginals() (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase method), 24  
 marginals() (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN method), 27  
 marginals() (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN method), 28  
 marginals() (snorkel.learning.gen\_learning.GenerativeModel method), 17  
 mark() (in module snorkel.learning.disc\_models.rnn.re\_rnn), 25  
 mark\_sentence() (in module snorkel.learning.disc\_models.rnn.re\_rnn), 25  
 Matcher (class in snorkel.matchers), 9  
 MentionScorer (class in snorkel.learning.utils), 29  
 MiscMatcher (class in snorkel.matchers), 9  
 ModelTester (class in snorkel.learning.utils), 30  
**N**  
 name (snorkel.learning.utils.ModelTester attribute), 30  
 NgramMatcher (class in snorkel.matchers), 9  
 Ngrams (class in snorkel.candidates), 8  
 NumberMatcher (class in snorkel.matchers), 9  
**O**  
 OPEN (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN attribute), 26  
 optional\_names (snorkel.learning.gen\_learning.GenerativeModel attribute), 17  
 OrganizationMatcher (class in snorkel.matchers), 9  
**P**  
 parse() (snorkel.parser.parser.URLParserConnection method), 5  
 ParserConnection (class in snorkel.parser.parser), 5  
 PersonMatcher (class in snorkel.matchers), 10  
 pid (snorkel.learning.utils.ModelTester attribute), 30  
 plot\_accuracy() (in module snorkel.learning.utils), 31  
 plot\_prediction\_probability() (in module snorkel.learning.utils), 31  
 post() (snorkel.parser.parser.URLParserConnection method), 5  
 Prediction (class in snorkel.models.annotation), 12  
 predictions() (snorkel.learning.classifier.Classifier method), 16  
 predictions() (snorkel.learning.disc\_learning.TFNoiseAwareModel method), 19  
 predictions() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression method), 21  
 predictions() (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression method), 23  
 predictions() (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN method), 26  
 predictions() (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase method), 24  
 predictions() (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN method), 27  
 predictions() (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN method), 28  
 predictions() (snorkel.learning.gen\_learning.GenerativeModel method), 17  
 PretaggedCandidateExtractor (class in snorkel.candidates), 8  
 PretaggedCandidateExtractorUDF (class in snorkel.candidates), 8  
 print\_scores() (in module snorkel.learning.utils), 31  
**R**  
 RandomSearch (class in snorkel.learning.utils), 30  
 RangeParameter (class in snorkel.learning.utils), 30  
 RegexMatch (class in snorkel.matchers), 10  
 RegexMatchEach (class in snorkel.matchers), 10  
 RegexMatchSpan (class in snorkel.matchers), 10  
 render() (snorkel.viewer.Viewer method), 33  
 representation (snorkel.learning.classifier.Classifier attribute), 16  
 representation (snorkel.learning.disc\_learning.TFNoiseAwareModel attribute), 20  
 representation (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression attribute), 21  
 representation (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression attribute), 23  
 representation (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN attribute), 26  
 representation (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase attribute), 24  
 representation (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN attribute), 27  
 representation (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN attribute), 28  
 representation (snorkel.learning.gen\_learning.GenerativeModel attribute), 17  
 reRNN (class in snorkel.learning.disc\_models.rnn.re\_rnn), 25  
 reshape\_marginals() (in module snorkel.learning.utils), 31  
 RNNBase (class in snorkel.learning.disc\_models.rnn.rnn\_base), 24  
 run() (snorkel.learning.utils.ModelTester method), 30  
**S**  
 save() (snorkel.learning.classifier.Classifier method), 16

save() (snorkel.learning.disc\_learning.TFNoiseAwareModelScorer (class in snorkel.learning.utils), 30 method), 20

save() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression (class in snorkel.learning.disc\_models.logistic\_regression), 29 method), 21

save() (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression (class in snorkel.learning.disc\_models.logistic\_regression), 23 method), 23

save() (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN (class in snorkel.learning.disc\_models.rnn.re\_rnn), 26 method), 26

save() (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase (class in snorkel.learning.disc\_models.rnn.rnn\_base), 24 method), 24

save() (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN (class in snorkel.learning.disc\_models.rnn.tag\_rnn), 27 method), 27

save() (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN (class in snorkel.learning.disc\_models.rnn.text\_rnn), 28 method), 28

save() (snorkel.learning.gen\_learning.GenerativeModel (class in snorkel.learning.gen\_learning), 17 method), 17

save\_marginals() (in module snorkel.annotations), 12

save\_marginals() (snorkel.learning.classifier.Classifier method), 16

save\_marginals() (snorkel.learning.disc\_learning.TFNoiseAwareModelScorer (class in snorkel.learning.utils), 20 method), 20

save\_marginals() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression (class in snorkel.learning.disc\_models.logistic\_regression), 21 method), 21

save\_marginals() (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression (class in snorkel.learning.disc\_models.logistic\_regression), 23 method), 23

save\_marginals() (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN (class in snorkel.learning.disc\_models.rnn.re\_rnn), 26 method), 26

save\_marginals() (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase (class in snorkel.learning.disc\_models.rnn.rnn\_base), 25 method), 25

save\_marginals() (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN (class in snorkel.learning.disc\_models.rnn.tag\_rnn), 27 method), 27

save\_marginals() (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN (class in snorkel.learning.disc\_models.rnn.text\_rnn), 28 method), 28

save\_marginals() (snorkel.learning.gen\_learning.GenerativeModel (class in snorkel.learning.gen\_learning), 17 method), 17

score() (snorkel.learning.classifier.Classifier method), 16

score() (snorkel.learning.disc\_learning.TFNoiseAwareModelScorer (class in snorkel.learning.utils), 20 method), 20

score() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression (class in snorkel.learning.disc\_models.logistic\_regression), 21 method), 21

score() (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression (class in snorkel.learning.disc\_models.logistic\_regression), 23 method), 23

score() (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN (class in snorkel.learning.disc\_models.rnn.re\_rnn), 26 method), 26

score() (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase (class in snorkel.learning.disc\_models.rnn.rnn\_base), 25 method), 25

score() (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN (class in snorkel.learning.disc\_models.rnn.tag\_rnn), 27 method), 27

score() (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN (class in snorkel.learning.disc\_models.rnn.text\_rnn), 28 method), 28

score() (snorkel.learning.gen\_learning.GenerativeModel (class in snorkel.learning.gen\_learning), 18 method), 18

score() (snorkel.learning.utils.MentionScorer method), 29

score() (snorkel.learning.utils.Scorer method), 30

search\_space() (snorkel.learning.utils.GridSearch (class in snorkel.learning.utils), 29 method), 29

search\_space() (snorkel.learning.utils.RandomSearch (class in snorkel.learning.utils), 30 method), 30

Sentence (class in snorkel.models.context), 3

SentenceNgramViewer (class in snorkel.viewer), 33

SlotFillMatch (class in snorkel.matchers), 10

snorkel.annotations (module), 12

snorkel.candidates (module), 8

snorkel.learning.classifier (module), 15

snorkel.learning.disc\_learning (module), 19

snorkel.learning.disc\_models.logistic\_regression (module), 21

snorkel.learning.disc\_models.rnn.re\_rnn (module), 25

snorkel.learning.disc\_models.rnn.rnn\_base (module), 24

snorkel.learning.disc\_models.rnn.tag\_rnn (module), 26

snorkel.learning.disc\_models.rnn.text\_rnn (module), 28

snorkel.learning.gen\_learning (module), 16

snorkel.learning.utils (module), 29

snorkel.If\_helpers (module), 33

snorkel.logistic\_regression (module), 9

snorkel.models.annotation (module), 11

snorkel.SparseLogisticRegression (module), 7

snorkel.models.context (module), 3

snorkel.parser.corpus\_parser (module), 4

snorkel.parser.doc\_preprocessors (module), 4

snorkel.parser.parser (module), 5

snorkel.viewer (module), 33

Span (class in snorkel.models.context), 3

sparse\_abs() (in module snorkel.learning.utils), 31

StableLabel (class in snorkel.models.annotation), 12

StableLogisticRegression (class in snorkel.learning.disc\_models.logistic\_regression), 22

split\_stable\_id() (in module snorkel.models.context), 4

start() (snorkel.learning.utils.ModelTester method), 30

summary\_score() (snorkel.learning.utils.MentionScorer method), 30

summary\_score() (snorkel.learning.utils.Scorer method), 30

TagRNN (class in snorkel.learning.disc\_models.rnn.tag\_rnn), 26

TemporaryContext (class in snorkel.models.context), 3

TemporarySpan (class in snorkel.models.context), 3

terminate() (snorkel.learning.utils.ModelTester method), 30

test\_LF() (in module snorkel.If\_helpers), 34

TextDocPreprocessor (class in snorkel.parser.doc\_preprocessors), 5

TextRNN (class in snorkel.learning.disc\_models.rnn.text\_rnn),  
28

TFNoiseAwareModel (class in  
snorkel.learning.disc\_learning), 19

TikaPreprocessor (class in  
snorkel.parser.doc\_preprocessors), 5

train() (snorkel.learning.disc\_learning.TFNoiseAwareModel  
method), 20

train() (snorkel.learning.disc\_models.logistic\_regression.LogisticRegression  
method), 22

train() (snorkel.learning.disc\_models.logistic\_regression.SparseLogisticRegression  
method), 23

train() (snorkel.learning.disc\_models.rnn.re\_rnn.reRNN  
method), 26

train() (snorkel.learning.disc\_models.rnn.rnn\_base.RNNBase  
method), 25

train() (snorkel.learning.disc\_models.rnn.tag\_rnn.TagRNN  
method), 27

train() (snorkel.learning.disc\_models.rnn.text\_rnn.TextRNN  
method), 29

train() (snorkel.learning.gen\_learning.GenerativeModel  
method), 18

training\_set\_summary\_stats() (in module  
snorkel.learning.utils), 31

TSVDocPreprocessor (class in  
snorkel.parser.doc\_preprocessors), 5

## U

Union (class in snorkel.matchers), 10

URLParserConnection (class in snorkel.parser.parser), 5

## V

Viewer (class in snorkel.viewer), 33

## W

word\_to\_char\_index() (snorkel.models.context.TemporarySpan  
method), 4

## X

XMLMultiDocPreprocessor (class in  
snorkel.parser.doc\_preprocessors), 5