

graphlab.nearest_neighbors.NearestNeighborsModel.query

`NearestNeighborsModel.query(dataset, label=None, k=5, radius=None, verbose=True)`

For each row of the input 'dataset', retrieve the nearest neighbors from the model's stored data. In general, the query dataset does not need to be the same as the reference data stored in the model, but if it is, the 'include_self_edges' parameter can be set to False to exclude results that match query points to themselves.

Parameters: **dataset** : SFrame

Query data. Must contain columns with the same names and types as the features used to train the model. Additional columns are allowed, but ignored. Please see the nearest neighbors `create()` documentation for more detail on allowable data types.

label : str, optional

Name of the query SFrame column with row labels. If 'label' is not specified, row numbers are used to identify query dataset rows in the output SFrame.

k : int, optional

Number of nearest neighbors to return from the reference set for each query observation. The default is 5 neighbors, but setting it to `None` will return all neighbors within `radius` of the query point.

radius : float, optional

Only neighbors whose distance to a query point is smaller than this value are returned. The default is `None`, in which case the `k` nearest neighbors are returned for each query point, regardless of distance.

verbose: bool, optional

If True, print progress updates and model details.

Returns: **out** : SFrame

An SFrame with the k-nearest neighbors of each query observation. The result contains four columns: the first is the label of the query observation, the second is the label of the nearby reference observation, the third is the distance between the query and reference observations, and the fourth is the rank of the reference observation among the query's k-nearest neighbors.

! See also

`similarity_graph`

- If both `k` and `radius` are set to `None`, each query point returns all of the reference set. If the reference dataset has n rows and the query dataset has m rows, the output is an SFrame with nm rows.
- For models created with the 'lsh' method, the query results may have fewer query labels than input query points. Because LSH is an approximate method, a query point may have fewer than 'k' neighbors. If LSH returns no neighbors at all for a query, the query point is omitted from the results.

Examples

First construct a toy SFrame and create a nearest neighbors model:

```
>>> sf = graphlab.SFrame({'label': range(3),
...                       'feature1': [0.98, 0.62, 0.11],
...                       'feature2': [0.69, 0.58, 0.36]})
>>> model = graphlab.nearest_neighbors.create(sf, 'label')
```

A new SFrame contains query observations with same schema as the reference SFrame. This SFrame is passed to the `query` method.

```
>>> queries = graphlab.SFrame({'label': range(3),
...                             'feature1': [0.05, 0.61, 0.99],
...                             'feature2': [0.06, 0.97, 0.86]})
>>> model.query(queries, 'label', k=2)
```

query_label	reference_label	distance	rank
0	2	0.305941170816	1
0	1	0.771556867638	2
1	1	0.390128184063	1
1	0	0.464004310325	2
2	0	0.170293863659	1
2	1	0.464004310325	2