

Instance-based Model: K-Nearest Neighbours

This directory contains the code used to generate our instance-based model for the competition. We decided to use a simple k-nearest neighbours (k-NN) classifier to represent our instance-based model. Please follow the instructions below to run the code.

About

k-NN is a non-parametric method used for classification and regression. In our case, we decided to use a k-NN classifier that would a tweet based on the k closest training examples in our feature space.

Instructions to run code

The instructions below will guide you in training and testing the model.

Training the model

To train the model, simply run the following command in MATLAB's command window:

```
KNN
```

This will run the training script, `KNN.m` for the model. The training script takes care of loading and preprocessing all of the training data. At the end, it will output the amount of time taken to train the model and the cost associated with testing the model on the training data.

Note: If you would like to save the trained model, uncomment line 19 in `KNN.m`. It will store the trained model as `KNN_model.mat`. Keep in mind that this will overwrite the model already provided in this directory.

Testing the model

To test the model, simply call the `test_knn_model` function. This function takes in 3 parameters:

1. The test bag of words (sparse matrix, `X_test_bag`)
2. The raw set of test tweets (cell matrix, `test_raw`)
3. The true test labels (`Y_test`)

`test_knn_model` will return the cost of predicting test data using our trained model. The function should be called in MATLAB's command window as the following:

```
test_knn_model(X_test_bag, test_raw, Y_test)
```

Note: Make sure your `X_test_bag` is a sparse matrix of size $N \times 10000$, where N denotes the number of samples in the data set. Also make sure `test_raw` is cell matrix of size $N \times 1$, and `Y_test` is double matrix of $N \times 1$ as well.

An example

Say we want to test our trained model on the training data. Make the following call in MATLAB's command window:

```
test_knn_model(X_train_bag, train_raw, Y_train)
```

This assumes that you have already loaded the training data into the workspace. If not, simply enter the following into MATLAB's command window:

```
load('../data/train.mat')
```