

# Discriminative Model: Logistic Regression

This directory contains the code used to generate our discriminative model for the competition. We decided to use a simple Logistic Regression classifier to represent this. Please follow the instructions below.

## About the model

Logistic Regression classifiers is a regression model where labels are categorical variables. This model assumes a form of  $P(Y|X)$ , which is then estimated from the training data.

## Instructions to run code

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

### Training the model,

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

```
logistic_regression
```

This will run the training script, `logistic_regression.m` for the model. The training script will take 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, general loss (error) of the trained model, and the cost associated with testing the model on the training data.

**Note.** This model was generated using the liblinear package, which has also been included.

### Testing the model

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

```
test_logistic_model(X_test_bag, test_raw, Y_test)
```

**Note:** Make sure your `X_test_bag` is a sparse matrix of size  $N \times 10000$ .