The four we implemented were a generative model, discriminative model, semi-supervised dimensionality reduction model, and an instance based method with the text data. We used a 10-fold cross validation to calculate the training error of the model.

**Generative Model**

Our generative model was a Naive Bayes Model. We used the Matlab function, fitcnb() and set the distribution within the model as multinomial distribution. The model had a training error of 0.80.

To run the model the Naive Bayes Model…

**Discriminative Model**

The discriminative model we fit was a cross-validated SVM classifier using Bayesian Optimization. The model classifies the tweet as 0 or 1 using fitcsvm() , and then uses bayesopt() to optimize the parameters from the SVM model with respect to the cross validation. The model had a training error of 0.81.

To run the discriminative model…

**Instance Based Method**

**Semi-Supervised Dimensionality Reduction**

To reduce the data, we reduced the dimensionality of the data with PCA using 150 principal components and a regularization value of 0.0001 with the pca() function. We then ran a Gaussian Mixture Model to model the reduced training data with two clusters.  with fitgmdist(). The training error was

To run the model..