CS534 Implementation Assignment 2: Naive Bayes

Amit Bawaskar, Michael Lam
EECS, Oregon State University

April 26, 2013

Abstract

In this assignment, we implemented the Naive Bayes classifier with the Bernoulli model and Multinomial model, and compared their performance.

1 Introduction

We implemented the Naive Bayes classifier to solve a document classification problem on the 20-newsgroup data set. Two models were implemented and compared for performance: Bernoulli model and Multinomial model.

2 Naive Bayes

Paragraph about Naive Bayes.

2.1 Model

Naive Bayes assumes that the features are independent.

2.2 Inference

Inference is performed by using Bayes Rule with the learned likelihood and prior probabilities, and using Decision Theory to select the class that maximizes the posterior probability.

2.3 Learning

Learning the likelihood probabilities for each feature and class is done by maximum a posteriori estimation, or equivalently applying Laplace smoothing to the maximum likelihood estimator here. Learning the prior probabilities for each class is done by finding the maximum likelihood estimator for it.

2.4 Implementation Details

Paragraph about implementation, namely to answer, "Please explain how you use the log of probability to perform classification." Also talk about using Laplace smoothing.

In this project we operated with the log of probabilities in order to avoid underflow issues. That is, for every multiplication and division operation, we instead used addition and subtraction of the log of the operands. We also stored the log of probabilities. For decision theory, we simply selected the class that maximizes the log of the posterior probability since the log function is a monotonically increasing function.

We also applied Laplace smoothing to the likelihood probability of each feature and class in order to assign a default prior to words that have not been encountered.

3 Bernoulli Model

Paragraph about Bernoulli model.

The overall test accuracy for the Bernoulli model is **0.772152**. Figure 1 shows the confusion matrix.

4 Multinomial Model

Paragraph about Multinomial model.

Report overall testing accuracy.

Report the confusion matrix.

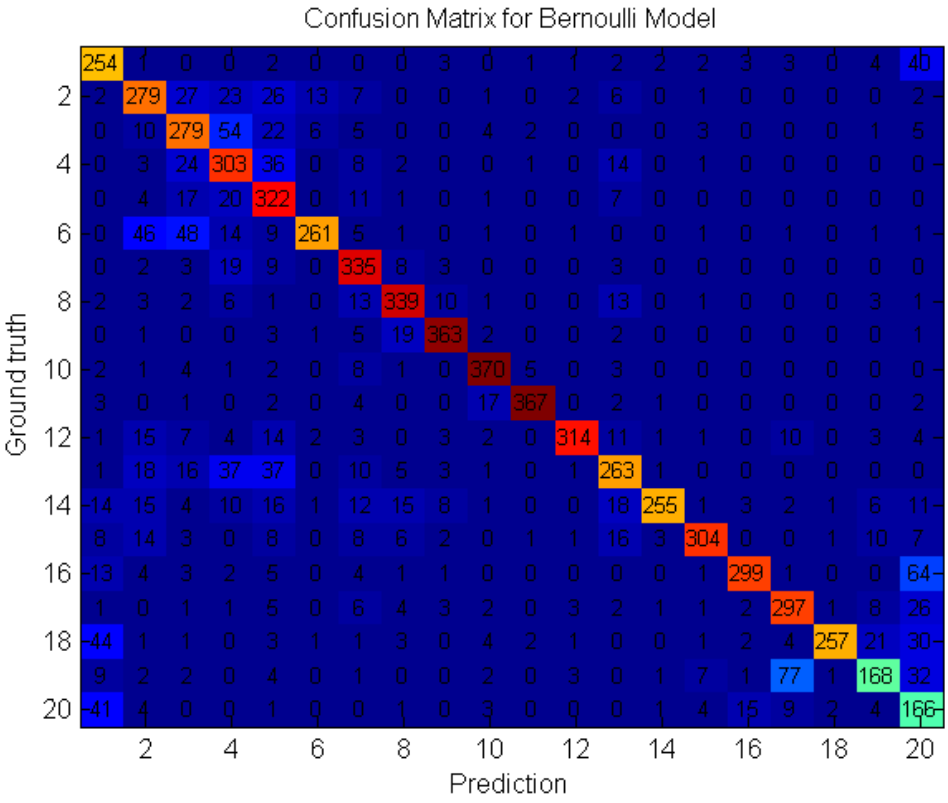

Figure 1: Confusion matrix for the Bernoulli model.