Vinay Gaba
UNI: vhg2105

COMS 4771 Machine Learning: Assignment 1

1) The problem consists of first computing the Maximum Likelihood Estimator for each feature separately and using this information in the Naïve Bayes Classifier to classify the given test data.

The **Naïve Bayes Classifier** is given by the following formula:

$$f*(x) = \underset{y \in Y}{arg\,max}\, Pr[Y = y] . Pr[X = x \mid Y = y]$$

where Pr [Y = y] is the probability of seeing label Y, also called as the class prior and Pr [X = x | Y = y] is the probability of value being x conditioned on the underlying label being y, also know as the class conditional distribution of X.

We then define the probability for any [y, x_1, x_2 . . . x_d] as

Pr[(X_1, X_2, . . . , X_d) = (x_1, x_2, . . . , x_d) | Y = y] = Pr [X_i = x_i | Y = y].

We will compute the **Maximum Likelihood Estimator** for each feature x = (x_1,……,x_{256}) where x ∈ X and X = {0,1}.

Maximum Likelihood Estimator for the Bernoulli Parameter ρ is given by:

$\rho_{ML} = \frac{1}{n} \Sigma\, x_i$ [where i = 1,…..n]

We will use this in the Naïve Bayes equation to compute the following values:

$$\Pr[Y = y] = \frac{\sum_{i=1}^{n}[[y^i = y]]}{n} = \frac{count(y)}{n} \; where\; y \in \{1, \ldots, k\}$$

In a similar fashion, we can compute the MLEs for the term $p_i(x_i/y)$ as follows:

$$\Pr[X = x \mid Y = y] = \frac{\sum_{j=1}^{n}[[y^j = y \; and \; x_i^{(j)} = x]]}{\sum_{j=1}^{n}[[y^j = y]]} = \frac{count_i(x \mid y)}{count\,(y)}$$

$$where\; y \in \{1, \ldots, k\} \; and \; x \in \{1, \ldots, d\} \; and$$

$$count_i(x \mid y) = \sum_{j=1}^{n} [[y^j = y \; and \; x_i^{(j)} = x]]$$

When we combine both the derivations, we get a Naïve Bayes Classifier with Bernoulli distribution given by the formula:

$$P(Y, X_1, X_2 \ldots\ldots, X_d) = P(Y = y) \prod_{i=1}^{d} P_i(X_i \mid Y)$$

$$where \; P(X_i \mid Y) = P(X_{i=}1 \mid Y)X_i + (1 - P(X_i = 0 \mid Y))(1 - X_i), X \in \{0,1\}, Y \in \{1..k\}$$

Number of errors in test data: 167

2) <u>Error Computed in the Naïve Bayes Classifier</u>

Error Type	Error
Training Error	0.1582
Test Error	0.1670

<u>Number of errors using k-nearest neighbor classifier with l2 norm</u>

k Value	Number of Errors
1	84
3	85
5	91

3) **<u>Error Computed for k-nearest neighbor classifier</u>**

k Value	Test Error	Training Error
1	0.0840	0
3	0.0850	0.0450
5	0.0910	0.0580

Citations

1. Prof. Satyan Kale's Slides
2. The Naive Bayes Model, Maximum-Likelihood Estimation, and the EM Algorithm by Prof. Michael Collins [http://www.cs.columbia.edu/~mcollins/em.pdf]

Study Group

Diksha Vanvari dhv2108@columbia.edu
Varun Shetty vs2567@columbia.edu