CS 464 Introduction to Machine Learning

Mustafa Göktan Güdükbay

21801740

1. The White Library

1.1. $S = \{1N, 2N, 1P, 2P, 3P, 1S, 2S, \ 3S\}$

1.2. $A = \{1N, 2P, 3P\}$

1.3.

I. The probability of an event is a real number and it must be greater or equal than zero and less than or equal to 1.
II. The sum of the probabilities of all possible events must be equal to 1.
III. If the tow events are mutually exclusive, the probability of either of two disjoint events occurring is the sum of probabilities of each event occurring separately.

1.4.

$$P(\{3S, 2N\}) = P(\{3S\}) + P(\{2N\}) = 0.045$$

$$P(\{3S, 2N, 2P\}) = P(\{3S\}) + P(\{2N\}) + P(\{2P\}) = 0.11$$

$$P(\{2P, 3S\}) = P(\{3S\}) + P(\{2N\}) = 0.06$$

We can find $P(\{2P\})$ by subtracting $P(\{3S, 2N\})$ from $P(\{3S, 2N, 2P\})$.

$$P(\{2P\}) = 0.065$$

We can find $P(\{2N\})$ by subtracting $P(\{2P, 3S\})$ from $P(\{3S, 2N, 2P\})$.

$$P(\{2N\}) = 0.05$$

We can find $P(\{3S\})$ by subtracting $P(\{2P\})$ from $P(\{2P, 3S\})$.

$$P(\{3S\}) = -0.005$$

$0.065 > 0.06$, which implies that $P(\{3S\}) = -0.005 < 0$. This is not possible according to axiom 1. I disagree with Donald. His estimations are not correct.

2. Cafe Customers

2.1. $P(Y < 3) = P(Y = 0) + P(Y = 1) + P(Y = 2)$

$= (C_0^{10})(0.7)^{10}(0.3)^0 + (C_1^{10})(0.7)^9(0.3)^1 + (C_2^{10})(0.7)^8(0.3)^2$

$= 0.7^{10} + 10 \times 0.7^9 \times 0.3^1 + 45 \times 0.7^8 \times 0.3^2$

$= 0.383$

2.2. $P(X = 2, Y = 0) = e^{-20} \times \frac{20^2}{2!} \times (0.7)^2$

$= 2.0199 \times 10^{-7}$

2.3. $E[Y] = 0.3N$

N depends on the random variable x, so the mean of x should be taken for N.

$E[Y] = 0.3 \times 20 = 6$

3. Spam Email Detection

3.1. The number of spam mails in the training set is 2911. The percentage of spam mails in the training set is 71.26%. The training set is skewed towards spam mails. I think having an imbalanced training set will affect the accuracy of the model because prior probabilities and likelihood probabilities depend on the percentage of classes in the training data. A possible solution would be using k-fold cross validation. The training and validation data together can be merged and divided into k parts. Then after each time, the training and validation data could be selected from different parts. By using this method, the model with the lowest average error percentage can be chosen.

3.2. Multinomial Naïve Bayes Model

Output:

```
Accuracy:  85.451197053407 %

Confusion Matrix

                 Actual
Classifier      Spam    Normal
     Spam        611      8
     Normal      150      317

Wrong Predictions:  158
```

3.3. Multinomial Naïve Bayes Model with Dirichlet Prior

Output:

```
Accuracy:  98.15837937384899 %

Confusion Matrix

                 Actual
Classifier      Spam    Normal
     Spam        751      10
     Normal      10      315

Wrong Predictions:  20
```

In Multinomial Naïve Bayes Model with Dirichlet, Prior the accuracy was higher compared to the Multinomial Naïve Bayes Model without Dirichlet Prior. Additionally, the number of false negatives was quite higher in the previous part compared to this model. In this model, the number of wrong predictions made decreased because in this model we assumed that each model existed at least once. In the previous model, the likelihood values were 0 for some cases. However, in this model, they were greater than zero. The logarithm of 0 in the previous model caused many false negatives. Dirichlet prior prevents probabilities from being zero so when we take the logarithm of such values they do not evaluate to negative infinity.

3.4. Bag-of-Words Representation and Bernoulli Naive Bayes Model

Output:

```
Accuracy:  83.60957642725599 %

Confusion Matrix

                 Actual
Classifier      Spam    Normal
     Spam        608      25
     Normal      153      300

Wrong Predictions:  178
```

In Bernoulli Naive Bayes Model the accuracy was lower compared to the other two Multinomial Naive Bayes Models. This model contained more wrong predictions because first of all, it did not take into account the fact that the number of occurrences of words can affect the accuracy. Besides, the Bernouilli Model did not use Laplace Smoothing. Multinomial Naive Bayes Model without Dirichlet Prior and Bernoulli Naive Bayes Model had similar true positives. The Multinomial Naive Bayes Model with Dirichlet Prior had more true positives compared to other models. However, the main difference between the models was the number of false negatives. The first and the last models had a higher number of false negatives compared to the second model. The second model had the highest accuracy.