Bengisu Buket Karadoğan
21802215 - Sec: 2
12.11.2023

CS464 Introduction to Machine Learning
Fall 2023
Homework 1

Question 1.1

$$P(X = two\ heads\ in\ row) = (1/2)((2/3)(1/2)^2 + (1/3)(1/4)^2) + (1/2)((1/2)(1/2)^2 + (1/2)(1/10)^2)$$
$$= 0.08333 + 0.01041 + 0.0625 + 0.0025$$
$$= 15874$$

Question 1.2

$$P(Y = fair \mid X) = \frac{0.0833 + 0.0625}{0.15874}$$
$$= 0.91867$$

Question 1.3

$$P(Y = red \mid X) = \frac{0.0025}{0.15874}$$
$$= 0.01574$$

Question 2.1

$$Likelihood = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2)$$

$$ln(Likelihood) = ln\left(\prod_{i=1}^{n} f(x_i; \mu, \sigma^2)\right)$$
$$= ln\left(\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2}\right)$$
$$= \sum_{i=1}^{n}\left[ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - (x_i - \mu)^2/2\sigma^2\right]$$
$$= -nln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{d}{d\mu}[ln(Likelihood)] = \frac{d}{d\mu}\left[-nln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2)(x_i - \mu)$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2)(x_i - \mu) = 0$$

$$\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\sum_{i=1}^{n}x_i - n\mu = 0$$

$$\mu_{MLE} = \frac{1}{n}\sum_{i=1}^{n}x_i$$

Question 2.2

$$Posterior \propto Likelihood \cdot Prior$$

$$Posterior \propto \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) \cdot \lambda e^{-\lambda\mu}$$

$$ln(Posterior) \propto ln\left[\prod_{i=1}^{n} f(x_i; \mu, \sigma^2) \cdot \lambda e^{-\lambda\mu}\right]$$

$$= -n ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 + ln\lambda - \lambda\mu$$

$$\frac{d}{d\mu}[ln(Likelihood)] = \frac{d}{d\mu}\left[-n ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 + ln\lambda - \lambda\mu\right]$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2)(x_i - \mu) - \lambda$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2)(x_i - \mu) - \lambda = 0$$

$$\sum_{i=1}^{n}\frac{(x_i - \mu)}{\sigma^2} - \lambda = 0$$

$$\sum_{i=1}^{n}(x_i - \mu) = \lambda\sigma^2$$

$$\sum_{i=1}^{n}x_i - n\mu = \lambda\sigma^2$$

$$\mu_{MAP} = \frac{\sum_{i=1}^{n}(x_i) - \lambda\sigma^2}{n}$$

Bengisu Buket Karadoğan
21802215 - Sec: 2
12.11.2023

Question 2.3

$f(x_i; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2}$

by choosing $\mu = 1$ and $\sigma = 1$

$f(x_i; 1, 1) = \frac{1}{\sqrt{2\pi}} e^{-(x_i-1)^2/2}$

The probability of the data point $x_{n+1} = 1$

$P(x_{n+1} \approx 1) = \int_{1-\epsilon}^{1+\epsilon} f(x_i; 1, 1)dx$

by choosing a small interval around $x_{n+1} = 1$

$P(0.9 \prec x_{n+1} \prec 1.1) = \int_{0.9}^{1.1} f(x_i; 1, 1)dx$

$= \int_{0.9}^{1.1} \frac{1}{\sqrt{2\pi}} e^{-(x_i-1)^2/2} dx$

The likelihood of the datapoint $x_{n+1} = 2$

The likelihood is the value of PDF at $x = 2$

$f(2; 1, 1) = \frac{1}{\sqrt{2\pi}} e^{-(2-1)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-1/2}$

Bengisu Buket Karadoğan
21802215 - Sec: 2
12.11.2023

Question 3.1.1

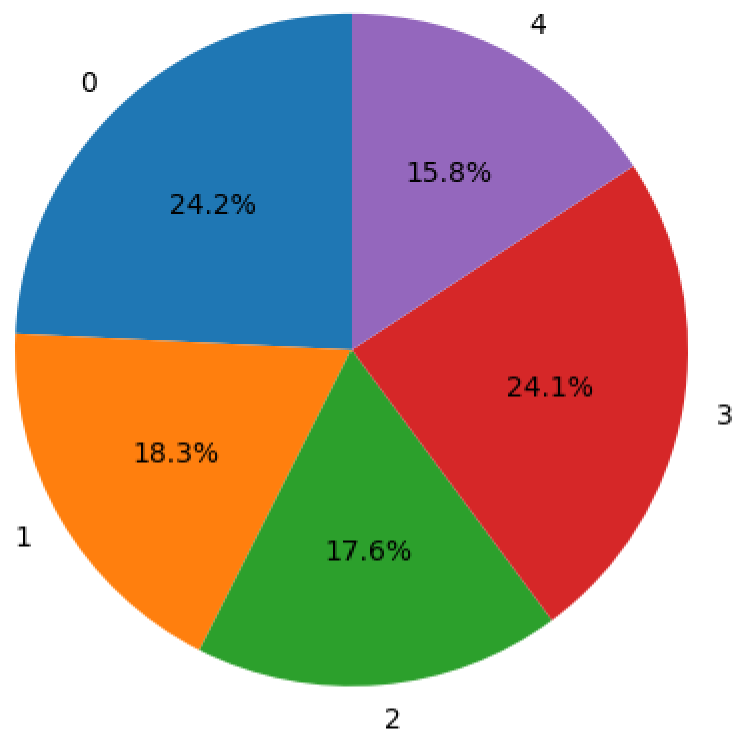

Figure 1: the percentages of each category in the train data

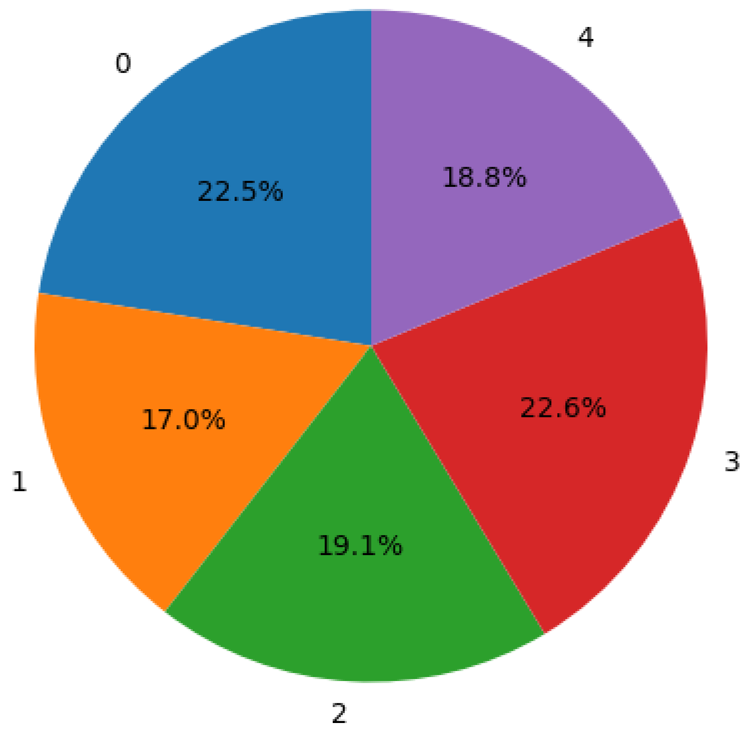

Figure 2: the percentages of each category in the test data

Bengisu Buket Karadoğan
21802215 - Sec: 2
12.11.2023

Question 3.1.2

Prior probabilities
$P(X = 0) = 0.22482014$
$P(X = 1) = 0.17026379$
$P(X = 2) = 0.191247$
$P(X = 3) = 0.22601918$
$P(X = 4) = 0.18764988$

Question 3.1.3

By looking at prior probabilities we can say that the training set is somewhat imbalanced since all classes do not have approximately the same probability. Having an imbalanced training set can affect the models as follows: bias toward the majority class, poor generalization, misleading accuracy, and difficulty in learning minority patterns.

Question 3.1.4

The word “alien” appears in the training documents 3 times.
The word “thunder” appears in the training documents 0 times.

$ln(P(alien \,|\, Y = Tech)) = ln(3/78433)$
$ln(P(thunder \,|\, Y = Tech)) = ln(0)$

Question 3.2

```
Accuracy: 0.242
Number of correct predictions: 135
Number of wrong predictions: 422
Confusion Matrix:
[[135.   0.   0.   0.   0.]
 [102.   0.   0.   0.   0.]
 [ 98.   0.   0.   0.   0.]
 [134.   0.   0.   0.   0.]
 [ 88.   0.   0.   0.   0.]]
```

Figure 3: the confusion matrix of Multinomial Naive Bayes with -inf values

```
Accuracy: 0.964
Number of correct predictions: 537
Number of wrong predictions: 20
Confusion Matrix:
[[129.   0.   3.   1.   2.]
 [  0.  94.   2.   0.   6.]
 [  1.   1.  94.   0.   2.]
 [  0.   0.   0. 134.   0.]
 [  0.   2.   0.   0.  86.]]
```

Figure 4: the confusion matrix of Multinomial Naive Bayes without -inf values

In order to avoid the number “$- inf$”, I change the probabilities equal 0 with 10^{-12}, an arbitrarily chosen small number. By making this adjustment, I ensure that my modal remains working even in the presence of 0 probabilities.

Question 3.3

```
Accuracy: 0.977
Number of correct predictions: 544
Number of wrong predictions: 13
Confusion Matrix:
[[131.   0.   2.   0.   2.]
 [  0.  97.   0.   0.   5.]
 [  1.   0.  96.   0.   1.]
 [  0.   0.   1. 133.   0.]
 [  1.   0.   0.   0.  87.]]
```

Figure 5: the confusion matrix of Multinomial Naive Bayes with Dirichlet prior

The Dirichlet prior affects the model as if it is a pseudocount. That is to say, the Dirichlet prior removes a small probability of observed events and distributes it to the unobserved events. As a result, the zero probability problem is prevented and the model gains robustness and generalization.

Bengisu Buket Karadoğan
21802215 - Sec: 2
12.11.2023

Question 3.4

```
Accuracy: 0.948
Number of correct predictions: 528
Number of wrong predictions: 29
Confusion Matrix:
[[128.   0.   4.   0.   3.]
 [  0.  82.   8.   0.  12.]
 [  0.   0.  98.   0.   0.]
 [  0.   0.   1. 133.   0.]
 [  1.   0.   0.   0.  87.]]
```

Figure 6: the confusion matrix of Bernoulli Naive Bayes

Thanks to additive smoothing terms (α) in the estimator function, Bernoulli Naive Bayes model performs better than Multinomial Naive Bayes without explicitly manipulating the 0 probabilities.

References

[1]J. Brownlee, “A gentle introduction to the bag-of-words model,” Machine Learning Mastery,
07-Aug-2019. [Online]. Available:
https://machinelearningmastery.com/gentle-introduction-bag-words-model/. [Accessed: 12-Nov-2023].

[2]S. Mazumder, “What is imbalanced data: Techniques to handle imbalanced data,” Analytics Vidhya,
23-Jun-2021. [Online]. Available:
https://www.analyticsvidhya.com/blog/2021/06/5-techniques-to-handle-imbalanced-data-for-a-clas
sification-problem/#:~:text=When%20we%20are%20using%20an,class%20which%20is%20called
%20undersampling. [Accessed: 12-Nov-2023].

[3]J. Brownlee, “A gentle introduction to imbalanced classification,” Machine Learning Mastery,
14-Jan-2020. [Online]. Available:
https://machinelearningmastery.com/what-is-imbalanced-classification/. [Accessed: 12-Nov-2023].

[4]H. Tripathi, “What is balance and imbalance dataset?,” Medium, 18-Jan-2021. [Online]. Available:
https://medium.com/analytics-vidhya/what-is-balance-and-imbalance-dataset-89e8d7f46bc5.
[Accessed: 12-Nov-2023].