Introduction to Machine Learning
CS464
Fall 2023

Homework 1

Deniz Aydemir
22001859

Question 1

$P(H^2) = Getting\ two\ heads\ in\ a\ row$
$P(B1) = Choosing\ Box\ 1$
$P(B2) = Choosing\ Box\ 2$
$P(Red) = Getting\ red\ coin$
$P(Blue) = Getting\ blue\ coin$
$P(Yellow) = Getting\ yellow\ coin$

Question 1.1

$$P(H^2) = P(B1 \cap Blue \cap H^2) + P(B1 \cap Yellow \cap H^2) + P(B2 \cap Blue \cap H^2) + P(B2 \cap Red \cap H^2)$$

$$\begin{aligned} = \quad & P(B1) * P(Blue \mid B1) * P(H^2 \mid Blue \cap B1) \\ & + P(B1) * P(Yellow \mid B1) * P(H^2 \mid Yellow \cap B1) \\ & + P(B2) * P(Blue \mid B2) * P(H^2 \mid Blue \cap B2) \\ & + P(B2) * P(Red \mid B2) * P(H^2 \mid Red \cap B2) \end{aligned}$$

$$\begin{aligned} = \quad & 1/2 * 2/3 * 1/4 \\ & + 1/2 * 1/3 * 1/16 \\ & + 1/2 * 1/2 * 1/4 \\ & + 1/2 * 1/2 * 1/100 \end{aligned}$$

$$= \quad 0.15875$$

Question 1.2

$$P(Blue \mid H^2) = P(Blue \cap H^2) / P(H^2)$$
$$= (P(B1 \cap Blue \cap H^2) + P(B2 \cap Blue \cap H^2)) / P(H^2)$$
$$= \frac{1/12 + 1/16}{0.15875}$$
$$= 0.91864$$

Question 1.3

$$P(Red \mid H^2) = P(Red \cap H^2) / P(H^2)$$
$$= (P(B1 \cap Red \cap H^2) + P(B2 \cap Red \cap H^2)) / P(H^2)$$
$$= \frac{0 + 1/400}{0.15875}$$
$$= 0.01575$$

Question 2

Let $X = x_1, x_2, ..., x_n$

Question 2.1

$$P(X \mid \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2 / 2\sigma^2}$$

$$ln[P(X \mid \mu, \sigma)] = - n * ln\sigma - \frac{n}{2} ln2\pi - \sum_{i=1}^{n} (x_i - \mu)^2 / 2\sigma^2$$

Since $\hat{\mu}_{mle} = argmax_{\mu} ln[P(X \mid \mu, \sigma)]$,

$$\frac{\delta}{\delta\mu} ln[P(X \mid \mu, \sigma)] = \sum_{i=1}^{n} (x_i - \mu) / \sigma^2 = 0$$
$$\Rightarrow \sum_{i=1}^{n} (x_i - \mu) = 0$$
$$\Rightarrow \sum_{i=1}^{n} x_i = n * \mu$$
$$\Rightarrow \hat{\mu}_{mle} = \frac{1}{n} * \sum_{i=1}^{n} x_i$$

Question 2.2

$$\hat{\mu}_{map} = argmax_{\mu} \, ln[\, P(X \mid \mu,\ \sigma) \ * \ \lambda e^{-\lambda\mu} \,]$$

Since $ln[\, P(X \mid \mu,\ \sigma) \ * \ \lambda e^{-\lambda\mu} \,] = lnP(X \mid \mu,\ \sigma) \ + \ ln\lambda e^{-\lambda\mu}$,

$$\frac{\delta}{\delta\mu} ln[\, P(X \mid \mu,\ \sigma) \ * \ \lambda e^{-\lambda\mu} \,] = \frac{\delta}{\delta\mu} lnP(X \mid \mu,\ \sigma) \ + \ \frac{\delta}{\delta\mu} ln\lambda e^{-\lambda\mu}$$

And we know that $\frac{\delta}{\delta\mu} lnP(X \mid \mu,\ \sigma) \ = \ \sum_{i=1}^{n} (x_i - \ \mu) \,/\, \sigma^2$ from Question 2.1. Therefore,

$$\begin{aligned}
\frac{\delta}{\delta\mu} ln[\, P(X \mid \mu,\ \sigma) \ * \ \lambda e^{-\lambda\mu} \,] \quad &= \frac{1}{\sigma^2} * \sum_{i=1}^{n} x_i \ - \frac{n*\mu}{\sigma^2} - \ \lambda = 0 \\
&\Rightarrow \sum_{i=1}^{n} x_i \ - n \ * \ \mu - \sigma^2 * \lambda \ = \ 0 \\
&\Rightarrow \sum_{i=1}^{n} x_i \ - \sigma^2 * \lambda \ = \ n \ * \ \mu \\
&\Rightarrow \hat{\mu}_{map} = \frac{1}{n} * \left(\sum_{i=1}^{n} x_i \ - \sigma^2 * \ \lambda\right)
\end{aligned}$$

Question 2.3

- The probability of a specific value in a continuous probability distribution, like the normal distribution, is technically zero. Thus, the probability that x_{n+1} is equal to 1 is 0.

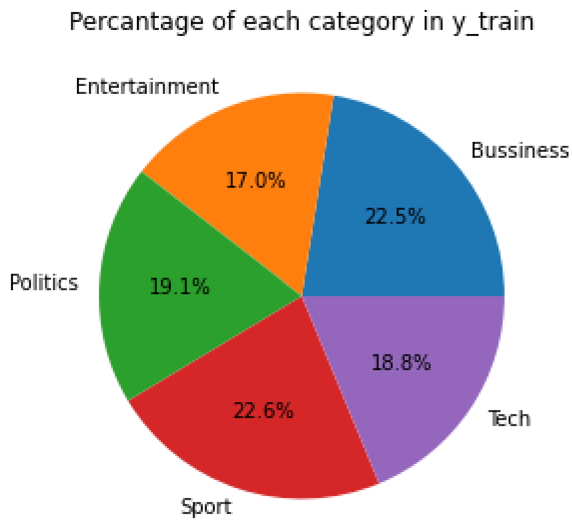
$$F(X \mid \mu \ = \ 1,\ \sigma \ = \ 1\,) \ = \frac{1}{\sqrt{2\pi}} e^{-(x-1)^2/2}$$

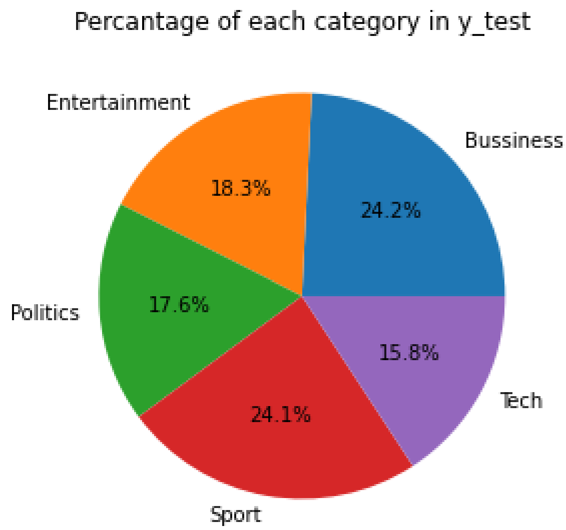
The likelihood of the data point $x_{n+1} \ = \ 2$ is

$$\begin{aligned}
F(X \ = \ 2 \mid \mu \ = \ 1,\ \sigma \ = \ 1\,) \quad &= \quad \frac{1}{\sqrt{2\pi}} e^{-1/2} \\
&= \quad 0.\,24197
\end{aligned}$$

Question 3

Question 3.1.1





Question 3.1.2

Prior probability for each class in the train set:

P(Bussiness) = 0.225
P(Entertainment) = 0.170
P(Politics) = 0.191
P(Sport) = 0.226
P(Tech) = 0.188

Prior probability for each class in the test set:

P(Bussiness) = 0.242
P(Entertainment) = 0.183
P(Politics) = 0.176
P(Sport) = 0.241
P(Tech) = 0.158

Prior probability for each class in the all data :

P(Bussiness) = 0.229
P(Entertainment) = 0.173
P(Politics) = 0.187
P(Sport) = 0.230
P(Tech) = 0.180

Question 3.1.3

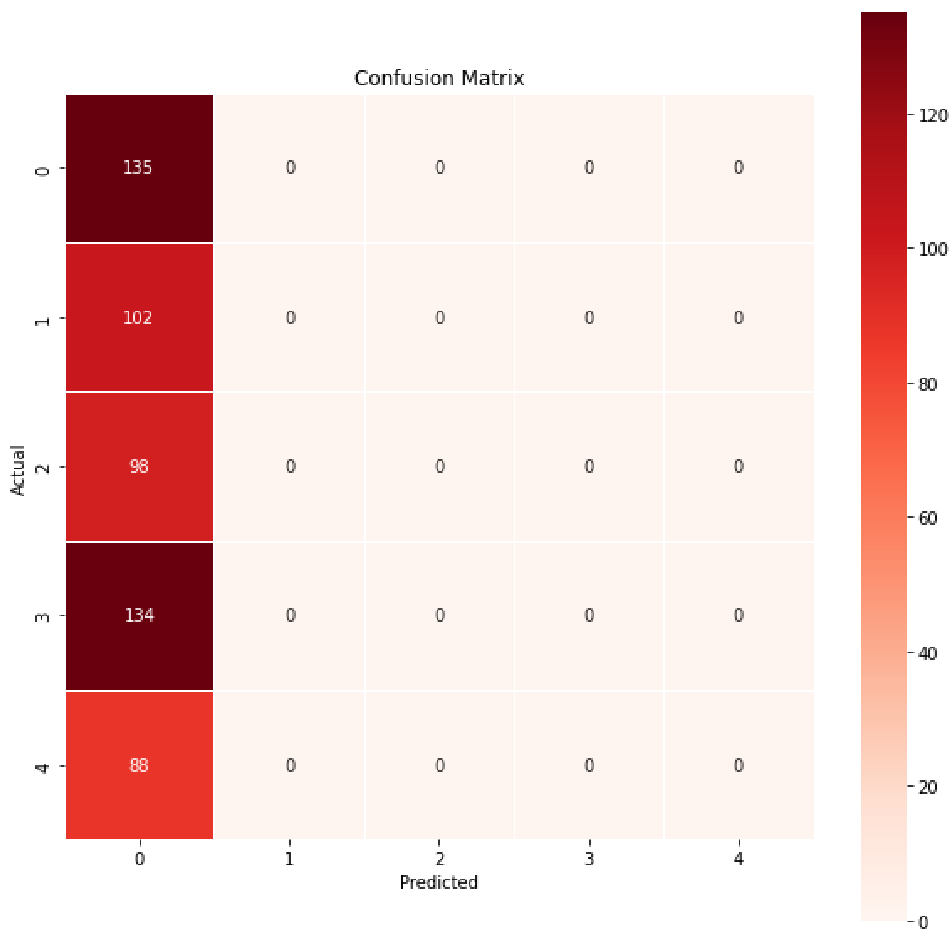
- I think the training set is balanced since the least frequent class (Entertainment) stands for 17.0% of the data whereas the most frequent class (Sport) stands for 22.6% of the data.
- Yes, in an imbalanced dataset, the model might become skewed towards the majority class since it sees more examples from this class. Model may perform poorly on the minority class while having good performance on the majority class. Moreover, the model may overfit the majority class, thus, perform poorly on the test set. Lastly, the model may have a high accuracy but low precision and recall.

Question 3.1.4

Alien count in Tech class in train set: 3
Thunder count in Tech class in train set: 0
Alien log ratio in Tech class in train set: -4.6476
Thunder log ratio in Tech class in train set: -inf

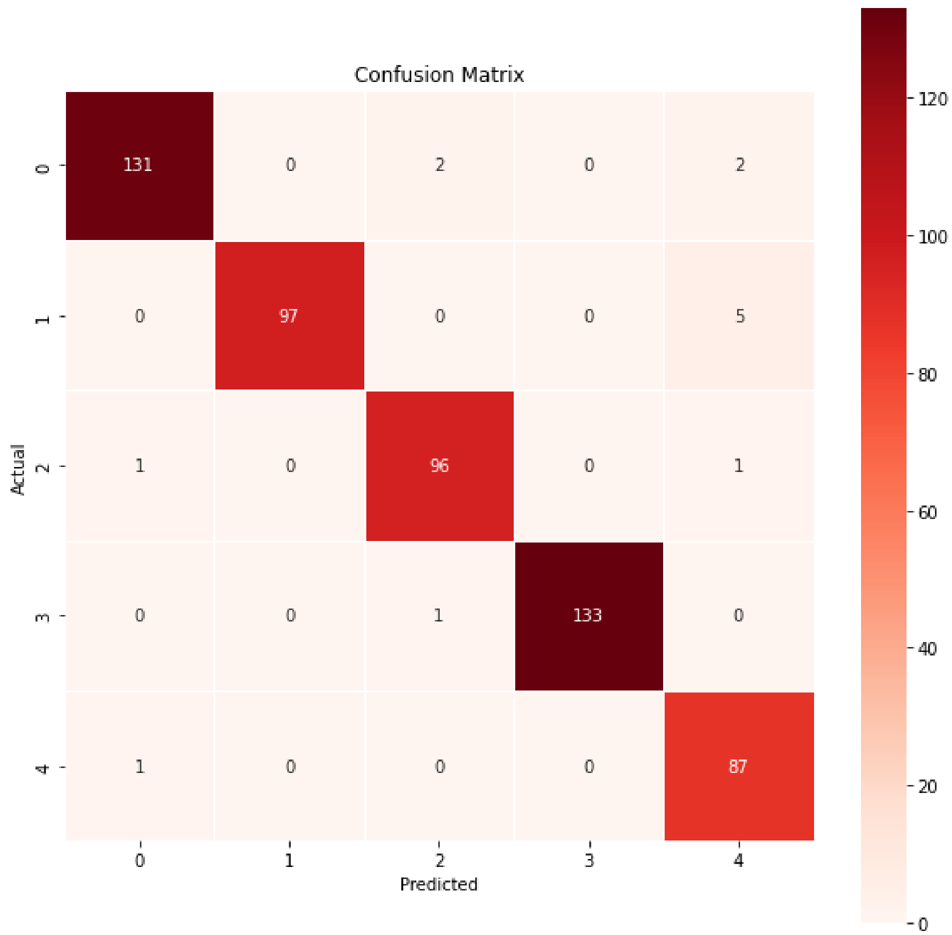
Question 3.2

Accuracy on test set when alpha is 0 (Multinomial): 0.242



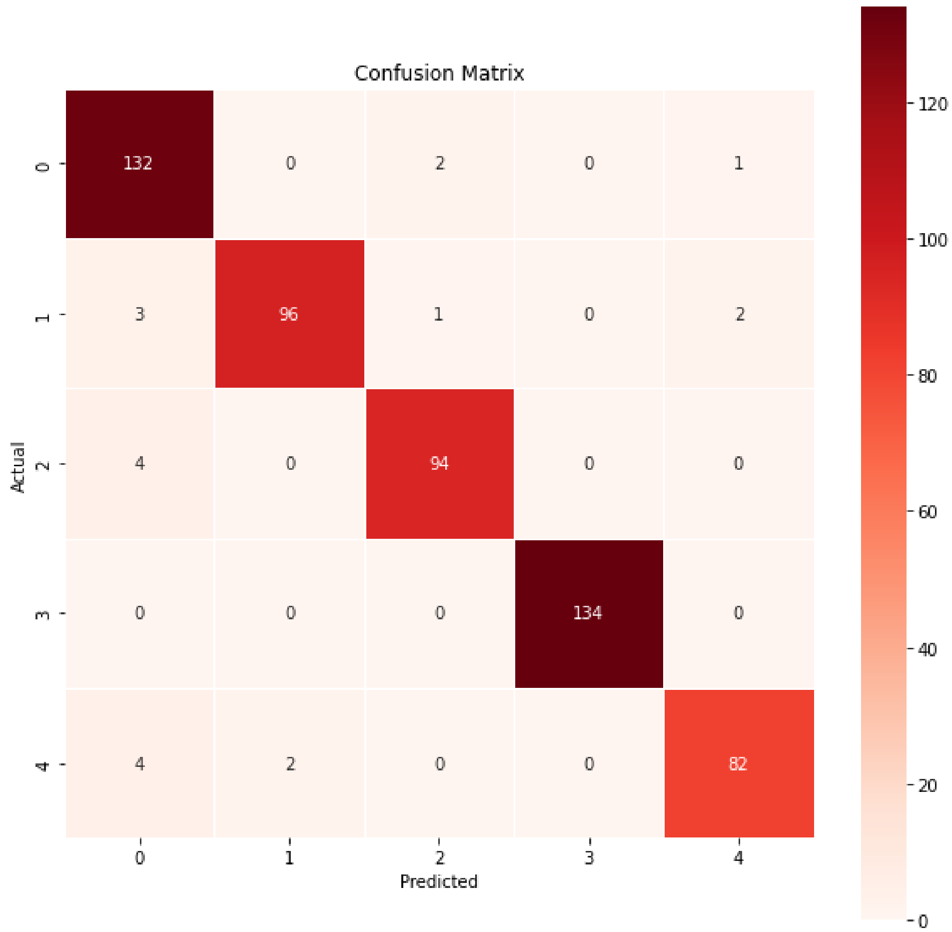
Question 3.3

Accuracy on test set when alpha is 1 (Multinomial): 0.977



Question 3.4

Accuracy on test set when alpha is 1 (Bernoulli): 0.966



- With the same alpha value ($\alpha = 1$) Bernoulli NBC is slightly worse (around 0.011 precision) than Multinomial NBC. This difference may result from the fact that we have relatively large data and long text. It seems that processing the information of multiple occurrences of a word (if any) puts Multinomial NBC one small step ahead of Bernoulli NBC when handling long texts.