CS464 Introduction to Machine Learning
Fall 2023
Homework 1

Q-1)

Defining Variables

Let X be the variable which defines the number of chosen box. X ∈ [1,2]

Let Y_i be the variable which defines the result of i'th toss. Y ∈ [H,T] for HEAD, TAIL

Let Z be the variable which defines the color of chosen coin. Z ∈ [R,Y,B] for RED, YELLOW AND BLUE RESPECTIVELY

Q-1.1)

$$P(Y1 = H, Y2 = H) \ = \ \Sigma x\, \Sigma z\, P(Y1 = H, Y2 = H, X, Z)$$

$$= P(Y1 = H, Y2 = H, X = 1, Z = B) + \ P(Y1 = H, Y2 = H, X = 2, Z = B) + \ P(Y1 = H, Y2 = H, X = 1, Z = Y) + \ P(Y1 = H, Y2 = H, X = 2, Z = R)$$

$$= \ P(Y2 = H|Y1 = H, X = 1, Z = B)P(Y1 = H|X = 1, Z = B)P(Z = B|X = 1)P(X = 1) +$$

$$P(Y2 = H|Y1 = H, X = 2, Z = B)P(Y1 = H|X = 2, Z = B)P(Z = B|\, X = 2)P(X = 2) +$$

$$P(Y2 = H|Y1 = H, X = 1, Z = Y)P(Y1 = H|X = 1, Z = Y)P(Z = Y|X = 1)P(X = 1) +$$

$$P(Y2 = H|Y1 = H, X = 2, Z = R)P(Y1 = H|X = 2, Z = R)P(Z = R|X = 2)P(X = 2)$$

$$= 1/2 * 1/2 * 2/3 * 1/2 +$$

$$1/4 * 1/4 * 1/3 * 1/2 +$$

$$1/2 * 1/2 * 1/2 * 1/2 +$$

$$1/10 * 1/10 * 1/2 * 1/2$$

$$= 1/12 + 1/96 + 1/16 + 1/400$$

$$= 0.15875$$

Q-1.2)

$$P(Z = B|Y1 = H, Y2 = H)$$

$$= \frac{P(Z=B, Y1=H, Y2=H)}{P(Y1=H, Y2=H)}$$

$$= P(Y1 = H, Y2 = H|Z = B)P(Z = B)$$

$$= \frac{P(Y2 = H|Y1 = H, Z = B)P(Y1 = H|Z = B)P(Z = B)}{P(Y1 = H, Y2 = H)}$$

$$= \frac{\frac{1}{2} * \frac{1}{2} * \frac{7}{12}}{0.15875}$$

$$= \frac{0.14583}{0.15875}$$

$$= 0.918635$$

Q-1.3)

The same way we can find the third question,

$$= \frac{P(Y2 = H|Y1 = H, Z = R)P(Y1 = H|Z = R)P(Z = R)}{P(Y1 = H, Y2 = H)}$$

$$= \frac{\frac{1}{10} * \frac{1}{10} * \frac{1}{4}}{0.15875}$$

$$= \frac{0.0025}{0.15875}$$

$$= 0.015748$$

We can also find for the Yellow coin,

$$\frac{P(Y2 = H|Y1 = H, Z = Y)P(Y1 = H|Z = Y)P(Z = Y)}{P(Y1 = H, Y2 = H)}$$

$$= \frac{\frac{1}{4} * \frac{1}{4} * \frac{1}{6}}{0.15875}$$

$$= \frac{0.010417}{0.15875}$$

$$= 0.0656168$$

To crosscheck, it adds up to 1

$$0.0656168 + 0.015748 + 0.91186 = 1{,}0002248$$

Q-2)

Q-2.1)

We need to find $MAX\ P(D|\mu) = \prod_1^n \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

For this, we need to take its derivative and equalize it to zero, but for the calculation to be easy we work with its ln function.

(Remark: Note that large operator ($\sum x$ and $\prod x$) are indexed for i which is a sub value x_i.)

$$P(D|\mu) = \ln(\prod_1^n \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}})$$

$$\frac{\partial P(D|\mu)}{\partial \mu} = \frac{\partial(\ln\left(\prod_1^n \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right))}{\partial \mu}$$

$$\frac{\partial\left(\sum_1^n \ln(\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}})\right)}{\partial \mu} = 0$$

$$\frac{\partial\left(\sum_1^n \ln\frac{1}{\sigma\sqrt{2\pi}} + \sum_1^n \ln(e^{-\frac{(x-\mu)^2}{2\sigma^2}})\right)}{\partial \mu} = 0$$

$$\frac{\partial\left(\sum_1^n \ln(e^{-\frac{(x-\mu)^2}{2\sigma^2}})\right)}{\partial \mu} = 0$$

$$\frac{\partial\left(\sum_1^n -\frac{(x-\mu)^2}{2\sigma^2})\right)}{\partial \mu} = 0$$

$$\sum_1^n \frac{(x-\mu)}{\sigma^2} = 0$$

$$\sum_1^n (x-\mu) = 0$$

$$n\bar{x} - n\mu = 0$$

$$\mu = \bar{x} = \frac{\sum_1^n x}{n}$$

Q-2.2)

Now we need to find $MAX\ P(\mu|D) = \frac{P(D|\mu)P(\mu)}{P(D)} = \frac{(\prod_1^n(\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}})\lambda e^{-\lambda x})}{P(D)}$

Because $P(D)$ does not depend on μ; we can just ignore that while doing the same procedure as Q-1.1)

$$MAX\ P(\mu|D) = \prod_{1}^{n}\left(\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)\lambda e^{-\lambda x}$$

$$MAX\ P(\mu|D) = \ln\left(\left(\prod_{1}^{n}\left(\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)\right)\lambda e^{-\lambda x}\ \right)$$

$$MAX\ P(\mu|D) = \sum_{1}^{n} ln\frac{1}{\sigma\sqrt{2\pi}} + ln\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) + \ ln\lambda e^{-\lambda x}$$

$$\frac{\partial P(\mu|D)}{\partial\mu} = \frac{\partial\left(\sum_{1}^{n} ln\frac{1}{\sigma\sqrt{2\pi}} + ln\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) + \ ln\lambda e^{-\lambda x}\right)}{\partial\mu}$$

$$\frac{\partial\left(\sum_{1}^{n} ln\frac{1}{\sigma\sqrt{2\pi}} + ln\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) + \ ln\lambda e^{-\lambda x}\right)}{\partial\mu} = 0$$

$$\frac{\partial\left(\sum_{1}^{n} ln\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) + \ ln\lambda e^{-\lambda x}\right)}{\partial\mu} = 0$$

$$\frac{\partial\left(\sum_{1}^{n} ln\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) + \ ln\lambda - \lambda x\right)}{\partial\mu} = 0$$

$$\left(\sum_{1}^{n}\frac{(x-\mu)}{\sigma^2} - \lambda\right) = 0$$

$$\mu = \frac{\sum_{1}^{n} x - \lambda\sigma^2}{n}$$

Q-2.3)

We can find it by putting the given numbers into the normal distribution function;

$$P(x_1, x_2, x_3 .. x_n|\sigma, \mu) = \prod_{1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$P(x_{i+1} = 1|\sigma = 1, \mu = 1) = \frac{1}{1\sqrt{2\pi}}\ e^{-\frac{(1-1)^2}{2*1^2}}$$

$$P(x_{i+1} = 1|\sigma = 1, \mu = 1) = \frac{1}{\sqrt{2\pi}}$$

$$P(x_{i+1} = 1|\sigma = 1, \mu = 1) = \ 0.39894$$

$$P(x_{i+1} = 2|\sigma = 1, \mu = 1) = \frac{1}{1\sqrt{2\pi}}\, e^{-\frac{(2-1)^2}{2*1^2}}$$

$$P(x_{i+1} = 1|\sigma = 1, \mu = 1) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}}$$

$$P(x_{i+1} = 1|\sigma = 1, \mu = 1) = 0.24197$$

Q3)Because I wrote my coding section as a report, I did not add the same things here, instead I made my projections and comments.

Overall Question 3 Projections and Comments

In this study, we explored three machine learning models for understanding news topics based on words: multinomial with no smoothing, multinomial with a smoothing operator of 1, and the Bernoulli model. We aimed to classify news into five categories: business, entertainment, politics, sport, and technology.

Firstly, the multinomial model without smoothing performed well, achieving an accuracy of 94.614%. Moving on to our second test with a smoothing operator of 1, the accuracy soared to 97.666%. Detailed confusion matrices were generated for both experiments to provide a deeper understanding of the models' performance.

Shifting to the third model, the Bernoulli model, we observed decent results but with the lowest accuracy of 91.382% among the three. Even with a smoothing operator, it struggled to match the performance of the multinomial models.

Beyond the numbers, it's essential to consider practical aspects like interpretability and computational efficiency. The multinomial models, both with and without smoothing, showed strong accuracy, making them promising for news topic classification. However, the choice between them depends on specific dataset characteristics and the balance between precision and computational complexity. Overall, this study sheds light on the strengths and considerations of different machine learning approaches in the context of news topic understanding.