CSE343/CSE543/ECE363/ECE563: Machine Learning Sec A (Monsoon 2023)
Quiz - 2

Date of Examination: 21.09.2023 Duration: 30 mins Total Marks: 15 marks

---

Instructions –

- Attempt all questions.
- MCQs have a single correct option.
- State any assumptions you have made clearly.
- Standard institute plagiarism policy holds.
- No evaluation without suitable justification.

---

0 marks if the option or justification of MCQs is incorrect.

1. For SVM, which options are correct- **[1.5 mark]**

   (A) Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane

   (B) Support vectors are data points that are far away from the hyperplane and influence the position and orientation of the hyperplane

   (C) Deleting the support vectors will change the position of the hyperplane

   (D) Deleting the support vectors won't change the position of the hyperplane

   (a) A and C.
   (b) B and C.
   (c) B and D.
   (d) A and D

   Solution: a) A and C , support vectors are used to find the optimal hyperplane and hence they directly affect the orientation of the hyperplane
   Rubric - binary marking - give marks if correct justification is given

2. You are given data of 1000 emails, of which 300 are spam and 700 are not spam. The dataset contains two features: the presence of the word "discount" and the presence of the word "lottery" in each email. You want to classify a new email that contains the words "discount" and "lottery" using Naive Bayes. Calculate the probability that the new email is spam. You know the following probabilities: **[3 mark]**

   1. $P(Discount|Spam) = 0.8$
   2. $P(Discount|NotSpam) = 0.2$
   3. $P(Lottery|Spam) = 0.6$
   4. $P(Lottery|NotSpam) = 0.1$

   Solution:

   To calculate the probability that the new email is spam using Naive Bayes, we can use Bayes' theorem:

$$P(S|D,L) = \frac{P(S)\cdot P(D|S)\cdot P(L|S)}{P(S)\cdot P(D|S)\cdot P(L|S) + P(NS)\cdot P(D|NS)\cdot P(L|NS)}$$

Where:

$$P(S) = \frac{300}{1000} = 0.3$$
$$P(D|S) = 0.8$$
$$P(L|S) = 0.6$$
$$P(NS) = \frac{700}{1000} = 0.7$$
$$P(D|NS) = 0.2$$
$$P(L|NS) = 0.1$$

Now, let's calculate the probability:

$$P(S|D,L) = \frac{0.3\cdot 0.8\cdot 0.6}{0.3\cdot 0.8\cdot 0.6 + 0.7\cdot 0.2\cdot 0.1} \approx 0.9114$$

Rubric - 2 marks for correct steps and 1 mark for correct answer. Give marks if just the numerator is calculated too. In that case the answer is 0.144

3. The total size of the dataset is 23. What is the number of experiments 't' that needs to be conducted given we are using the leave one out cross-validation technique? **[1 mark]**

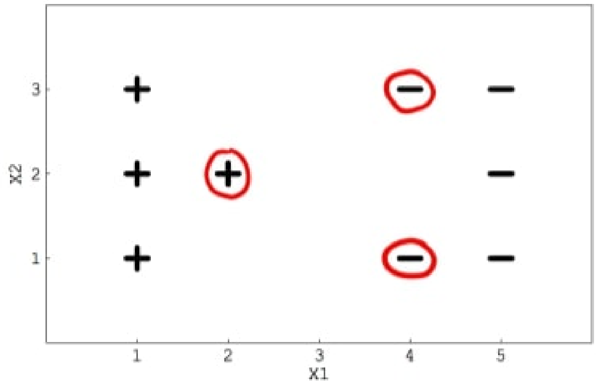
   1. 1
   2. 22
   3. 23
   4. 24

   Solution: c) 23, each sample will be used for validation and the model will be trained on rest 22 samples. This will be done for all the samples (that is taken for validation), thus 23 experiments will be conducted. Rubric - binary marking - give marks if correct justification is given

4. Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that represent support vectors. Will the decision boundary change if we: **[1.5 mark]**

   (A) Remove (2,2) from our dataset
   (B) Remove (5,1) from our dataset
   (C) Add (3,2) point to the plus class
   (D) Add (1.5,1.5) to the plot

   (a) A and B.
   (b) A and C.
   (c) A, D and C.
   (d) None



Attribute	Movie Lover		Movie Hater	
	Mean	Standard Deviation	Mean	Standard Deviation
Popcorn	24 g	4 g	8 g	2 g
Candy	20 g	5 g	50 g	8 g

Solution: b) A and C
Removing (2,2) changes the support vector for the plus class, hence the decision boundary will change. Adding (3,2) to the plus class has a similar effect. Removing (5,1) doesn't affect the support vectors, and hence the OSH remains the same. Adding (3/2,3/2) to the plus class doesn't change the OSH. It was announced that the point would be added to the plus class.
Rubric - binary marking - give marks if correct justification is given

5. Which type of Naive Bayes classifier is most appropriate for a dataset with binary features, such as presence or absence of specific attributes? **[1 mark]**

    1. Multinomial Naive Bayes
    2. Gaussian Naive Bayes
    3. Bernoulli Naive Bayes
    4. Poisson Naive Bayes

Solution: c) Bernoulli Naive Bayes

The Bernoulli Naive Bayes classifier is specifically designed for binary feature data. It models the distribution of each feature as a Bernoulli distribution (which is a discrete distribution with two possible outcomes, typically 0 and 1). It's commonly used when dealing with text data where each feature represents the presence or absence of a word in a document .
Rubric - binary marking - give marks if correct justification is given

6. Gaussian Naive Bayes: Consider a scenario where we classify whether a person will love or hate a movie based on the amount of popcorn (g) and candy (g) they ate. Based on prior knowledge, it is known that 50 % of people love movies in general. Now, it is found that the distribution of popcorn and candy consumed by movie lovers and haters are Gaussians with the following parameters. Now, suppose a new customer buys 20g of popcorn and 25g of candy. Use Gaussian Naive Bayes to predict whether he loves or hates the movie refer table of Attribute movie . - **[4 mark]**

Solution:

$$P(X = x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where:

$$\begin{aligned} P(x) &= \text{PDF of the Gaussian distribution} \\ x &= \text{Random variable} \\ \mu &= \text{Mean of the distribution} \\ \sigma &= \text{Standard deviation of the distribution} \end{aligned}$$

$$P(\text{loves}|\text{popcorn} = 20, \text{candy} = 25) = P(\text{loves}) \cdot P(\text{popcorn} = 20|\mu = 24, \sigma = 4) \cdot P(\text{candy} = 25|\mu = 20, \sigma = 5)$$
$$= 0.5 \cdot 0.061 \cdot 0.048 = 0.0015 \quad (1.5 marks)$$

$$P(\text{hates}|\text{popcorn} = 20, \text{candy} = 25) = P(\text{hates}) \cdot P(\text{popcorn} = 20|\mu = 8, \sigma = 2) \cdot P(\text{candy} = 25|\mu = 50, \sigma = 8)$$

$$= 0.5 \cdot 3.038 \cdot 10^{-9} \cdot 0.00038 = 5.77 \cdot 10^{-13} \quad (1.5 marks)$$

Thus, the customer will likely love the movie (1 mark).

7. What is the difference between Logistic Regression and SVM **[3 marks]**
Solution:

1. SVM tries to finds the "best" margin (distance between the line and the support vectors) that separates the classes and this reduces the risk of error on the data, while logistic regression does not, instead it can have different decision boundaries with different weights that are near the optimal point.
2. SVM is based on geometrical properties of the data while logistic regression is based on statistical approaches.
3. The risk of overfitting is less in SVM, while Logistic regression is vulnerable to overfitting.
4. In SVM, we directly classify the data point by calculating W^TX while logistic regression requires an additional sigmoid activation and the classification is done based on the probability score.

.
Rubric - min 3 points, 1 mark per correct point