AI 511 – Machine Learning

Tutorial 2

September 10, 2021

Instructions

- Question 1 & 2 is a the question given by sir in class.

- Question 3 is a **Subjective problem**. Please do so using pen/paper, tablets, or Latex.

- Questions 4 to 7 are **Multiple Choice Questions**. More than one option can also be correct. Make sure you mark them all.

- **Submission Details**

    - Deadline: **September 13th, 9:00 AM**
    - Format: **PDF** through Direct Message on **Slack**.

Subjective

1. Derive the parameters μ and σ using Maximum Likelihood Estimation for normal distribution(Univariate Gaussian).

$$P(X = x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

[Skip in case you have submitted it earlier]

2. For a linear regression with multiple features, called multivariate linear regression, give the update step in case of gradient descent. η is the learning rate. Do this for 3 parameters following the equation $y = w^{(0)} + w^{(1)}x_1 + w^{(2)}x_2$. Try to do this in vectors.

3. Derive the value of the parameter θ using Maximum Likelihood Estimation for the following.

$$P(X = x) = \theta \, x_o^\theta \, x^{-\theta-1} \qquad [x_o \text{ is a constant}]$$

Trivia: This probability density function called Pareto distribution which is a density function sometimes used in tasks related to economics.

Multiple Choice Questions

4. If you are given a data set of 1000 points and have been asked to predict house prices based on few features like the size of the house, location, age of the house, whether there is a metro station nearby, etc, and you have decided to use a linear regression model for the task. Then which of the following ways would be good to evaluate your model before deploying it?

    A. Just give up on the task

    B. Train on given 1000 data points and test it outside directly. This is the only way in which you would have maximum data points to train your model.

    C. Split the data into 2 parts, 800 for testing, 200 for training. In this way, you would have enough data to test before deploying it.

    D. Split the data into 2 parts, 800 for training, 200 for testing. In this way, you would have enough data to train the model before deploying it

5. We know that applying logarithm on the cost function makes our lives simpler. However, why you do think that it is even valid to apply logarithm? Because Logarithmic function

    A. is Linear

    B. is Differentiable

    C. is Monotonic

    D. cannot handle negative values

6. What are the assumptions taken for a Naive Bayes Classifier?

    A. Features of individual data points are independent of each other

    B. All the data points are uniformly distributed

    C. Each data point is independent of each other

    D. There exist at least 2 different target classes

7. Consider sigmoid function,

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

You have decided you want to use the sigmoid function for classification of data. If $\sigma(x)$ represents that probability of data point x belongs to class 1, determine a function, say $f(x)$, such that the output would be discrete i.e., either 1 or 0? Say you want to map it to either class 1 or class 0.

    A. $f(x) = \lfloor (\sigma(x)) \rfloor$ (Floor)

    B. $f(x) = \lceil (\sigma(x)) \rceil$ (Ceil)

    C. $f(x) = \text{Round}(\sigma(x))$

    D. $f(x) = \sigma(x)$

. . .