CSAL4243
Introduction to Machine Learning

Quiz 2

Each question carry same marks. Consider all models are linear regression with one variable and mean squared error as cost function.

1. Predicting semester GPA of a student based on his previous academic record is a classification problem. Yes/No and why ?

Ans: GPA is a real number, hence it's a regression problem.
Note: If someone says that its a classification problem and argue that number of possible grades combination is finite then she is right too. Although that would be a lot of classes.

2. What are the situations in which kNN is not a good choice to use ?

Ans: When dataset has very large number of features or very large number of samples.

3. In logistic regression, a point x on the decision boundary results in a predicted value h(x) of 0. Yes/No and why ?

Ans: No, its 0.5 as logistic function $h(x) = \frac{1}{1+e^{-\theta^T x}}$ returns 0.5 when $\theta^T x = 0$ which happens on decision boundary.

4. Provide the following information from assignment 1. Linear Regression.
    a. Dataset name: House Price prediction
    b. Output variable y: Sale Price
    c. Name any three features used in assignment X:
    Check features at
    https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data

5. In real world problems we will never get a cost of ∞ in logistic regression. Yes/No and why?

Ans: Yes, because cost is ∞ when $h(x) = \frac{1}{1+e^{-\theta^T x}}$ is either 0 or 1 which only happens when $\theta^T x = \pm\infty$ which only happens when a point is at infinity from decision boundary.