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. Predict how many A's a student will take in a semester based on his previous academic record is a classification problem. Yes/No and why ?

Ans: Number of A's is a classification problem since we are classifying among few distinct classes i.e. 0,1,2,3,4,5,6.

2. In kNN, what would happen if K is very large i.e. close to number of samples in the dataset ?

Ans: Most of the points will be classified as the class having highest number of elements.

3. 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

4. In logistic regression, a point x on the decision boundary results in a predicted value h(x) of 1. 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.

5. Logistic regression can not be used for non-linear decision boundaries. Yes/No and why?

Ans: It can be used for non-linear boundaries. You have to add high order polynomials to X.