Data Science Workshop-2 (CSE 2196)

ASSIGNMENT-4(LOGISTIC REGRESSION, KNN)

1. Write down the prediction function and cost function and the corresponding python code in the context for logistic regression.
2. Define types of logistic regression.
3. List the difference between linear regression and logistic regression.
4. Let you have given the following dataset:

x_1	x_2	y
0.5	1	0
1	2	0
1.5	2.5	1
2	3	1

   where x_1, x_2 are independent variable and y is dependent variable. In the context of logistic regression, find the optimized parameters after 3rd iteration. Find prediction for [1,1.5] w.r.t. the optimized parameter.
5. Explain K-Nearest Neighbor (KNN) algorithm.
6. How do you choose the optimal k for KNN model?
7. Suppose you have given the following dataset:

x_1	x_2	y
0.5	0.5	0
0.5	1	0
1	1	0
2	2.5	1
2.5	3	1
3	3	1

   Where x_1, x_2 are independent variable and y is dependent variable. Predict the class for [1.5,1] for k=2,3 respectively.
8. If the dataset is imbalance, then can the prediction by KNN be bias? Explain with an example.