CS 414 – Artificial Intelligence
Assignment No 2

Due on Monday, 11th December

Download a data set from Kaggle/ UCI Machine learning Repository/ https://datasetsearch.research.google.com/ related to classification or regression. Select the data not later than 3rd Dec and inform your CR to make sure no more than one student is solving the same problem. Make sure that data set should contain at least 4 features. CR to inform me list of selected problem in the following format by 4th Dec.

After selecting the problem and downloading the data set, perform the following

1. Load the dataset.
2. Perform necessary preprocessing.
3. Train/use following models on the data sets using all the features.
    a. Neural Network with one hidden layer (use number of hidden units ranging from 2-5)
    b. K-Nearest Neighbor (use appropriate distance metric and value of K to show your results)
4. Train NN models by initializing weights in two different ways (i) initialize weight vector with zero values (ii) initialize weight vector with small random values (for model 1 and 2 only). See the effect of training by keeping the learning rate fix at 0.01. Describe which weight initialization method is better and why in your report using iterations vs loss graphs.
5. Once you are done with the finalizing which weight initialization method is more appropriate. Run your models (1 and 2) using different learning rates 0.1,0.05,0.01,0.005,0.001 and show comparison using training graphs.
6. Perform the experiments by dividing your data sets into three different sets namely training set, validation set and test set. Demonstrate how your model is avoiding overfitting using early stopping.
7. Validate your models using test sets and report their testing and training accuracies.
8. Write a short report describing the problem, dataset and your produced results along with code and comments.

Reg No	Name	Dataset selected (along with 2–3-line description of the problem)