MIDTERM EXAM

SET 1 :(For odd machine number)

Linear regression

    a. Download weather.csv and create a new data set with attributes < MaxTemp, Rainfall, Humidity3pm, Cloud3pm, RainToday>. split 80% of the data to training set while 20% of the data to test. Load the training set.

    b. Implement gradient descent algorithm. Use the error function as the logarithm of hyperbolic cosine as shown in below equation

$$L(y, y^p) = \sum_{i=1}^{n} \log(\cosh(y_i^p - y_i))$$

    Where yi is the actual output and yi^p is the predicted output. [Hint: Log cosh error is a convex function and its gradient is tanh]

    c. Compute the prediction error and plot the error curve.
    d. Compare the prediction error with the sci-kit learn linear regression function error

SET 2 : (For even machine number)
    Logistic regression

    a. Download the training and test dataset to classify the income of a person as high or low. Training and test datasets are given in different files.
    b. Plot relationship between the attributes, 'age', capital-loss, capital-gain, and education-num
    c. Extract numeric features from both the training and test dataset.
    d. Implement gradient descent algorithm for logistic regression. The cost function should be taken as follows.

$$J(\theta) = \frac{1}{2m}[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^i)^2 + \lambda \sum_{j=1}^{n} \theta j^2]$$

    Where λ = 0.6

    e. Show the confusion matrix.
    f. Compare with the inbuilt logistic function.