Assignments (CS551 DL 2021)

Q1. Find out the eigen value and eigen vector of a given matrix using python library
X = [[1, 2, 3],
[2, 3, 4],
[4, 5, 6]]

Q2. Find the dot product of two vectors v = [1,2] and w =[1,1]

Q3. Find out the trace of a matrix given below using python.
X = [[1, 2, 3],
[4, 5, 6],
[7, 8, 9]]

Q2: Write a python program to perform the following set of operations on the augmented matrix derived from the system of linear equations given below.

(a) Generate Row Echelon Form (**REF**)
(b) Generate Reduced Row Echelon Form (**RREF**)
(c) Perform **Gaussian elimination** to solve the below system of linear equation
(d) Perform **Gauss–Jordan elimination** to solve the below system of linear equation

Test Case:
Input:
X+y+2z = 9
2x+4y-3z = 1
3x+6y-5z =0
Output:
x = 1, y =2 , z =3

Matrix factorization using SGD

Q2. Assume R is a mXn matrix with each entry is an integer in [0, 5] i.e. moving rating m users have given to n movies. Find the factorization of matrix $R = P^T Q$ using approximation method.
Where P and Q have size ($k \times m$) and ($k \times n$) and k<m and k<n.
Update rule for Pc and Qc (column vectors of P, Q respectively) are as follows:

Pc<- Pc +s*(eQc-t*Pc)
Qc<- Qc +s*(ePc-t*Qc)

Where s, t are from [0,1]. e= R_{i, j} - $(Pc)^T Qc$.

$E = ||R - P^T Q||_F$

1)Find k, s and t for least E. (use hyperopt or sklearn library for gridsearch.)
2)Plot graph for E vs k.

*Each user (Pc) and movie (Qc) are defined by k features (k sized 1-d vector) which best describes their characteristics in latent space.

Q3. Find the partial derivative of $||LU - A||_F$ (a function of squared frobenius norm) with respect to U. Where L, U and A are all matrices.

Linear Regression

Link: https://raw.githubusercontent.com/manishbhnau/Repo/master/Advertising.csv

Q1. Find the linear regression model which best predicts the dependent variable ("sales") on validation dataset.

- Divide data in 80-20 % for training and testing set.
- Use 10 fold cross validation set (10% of training dataset).

Find regression co-efficients and intercepts. Report confusion matrix on test dataset.

Q2. This question involves the use of multiple linear regression on the Auto data set (https://www.kaggle.com/uciml/autompg-dataset).
Compute the matrix of correlations between the variables

- Fit a least square linear model, You will need to exclude the name variable which is qualitative.
- Use the multiple linear regression with mpg as the response and all other variables except name as the predictors.

Logistic Regression:
Q5. The objective is to build a classifier that can predict whether an application will be admitted to the university (class 0) or not (class 1).

Dataset details :The data consists of marks of two exams for 100 applicants. The target value takes on binary values 1,0. 1 means the applicant was admitted to the university whereas 0 means the applicant didn't get an admission. Download data from the following link.
Dataset link:
https://github.com/animesh-agarwal/Machine-Learning/blob/master/LogisticRegression/data/marks.txt

(a) Calculate precision, recall, accuracy and f1 score.
(b) Visualize the confusion matrix using Heatmap.