Applications of Engineering Design

Dr. J. Hossain

Due Date: Aug 15, 2024 – 11:59pm

Instructions

1. Read problems carefully
2. You can use python for your calculation if needed.

P-1: In class you have learnt how to design linear regression model give data points. Given 15 students' hours spent in lectures, home works, and corresponding marks as

Hrs_Lec =[15, 18, 33, 22, 5, 29, 8, 5, 30, 8, 16, 1, 24 20, 2],

Hrs_HW =[7, 10, 7, 6, 8, 10, 6, 8, 4, 10, 6, 5, 10, 5, 2] and

Marks =[57 66 88 64 48 100 49 46 77 50 64 29 86 59 26].

In Lecture, you have learnt how to design the parameters of the linear regression that minimizes the mean square error (MSE). You know the solution for the optimal parameters is $\hat{\boldsymbol{w}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$.

1) Write **X** and **y** for this linear regression problem Marks **(1+1)**.
2) Write a python code to calculate optimal parameters using given equations. (Marks: 3)

P-2 : Concrete is the most important material in civil engineering. The concrete compressive strength is a function of age and ingredients. You want to build a regression model to predict the compressive strength based on age and ingredients. The relevant data set can be downloaded from https://www.kaggle.com/datasets/ryanholbrook/dl-course-data?select=concrete.csv

1) Make a Pandas DataFrame by reading the .csv file and print five rows of the DataFrame from the top. **Mark: 1**
2) How many features (N) do you have for this data set? **Mark 0.5**
3) Write down the names of all the features. **Mark 0.5**
4) Is your problem a supervised learning problem or an unsupervised learning problem? **Mark 0.5**
5) Plot feature maps by considering one feature at a time (you can plot feature map for any three features). **Mark 1**
6) In Lecture you have learnt how to design/optimize the parameters of the linear regression model that minimizes the mean square error (MSE). You want to design a linear regression model with only one input feature (you can consider any of the input features). Python is code is given below.
    a. Report the test MSE of your linear regression model. **Mark 1**

Python Code for P-2

```
import pandas as pd
from sklearn.model_selection import train_test_split # To split data set into training and testing
from sklearn.linear_model import LinearRegression # To use LinearRegression Model
from sklearn.metrics import mean_squared_error # to calculate MSE
import matplotlib.pyplot as  plt



# Load and prepare the dataset
data = pd.read_csv('concrete_data.csv')


X = data[['water']]  # Ensure X is a DataFrame
y = data['concrete_compressive_strength']  # Ensure y is a Series


# Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42) #
random_state =starting state. you will have same test/training set for each run

Train_set_size=800
X_train_set=X_train.iloc[0:Train_set_size]
y_train_set=y_train.iloc[0:Train_set_size]

linear_model=LinearRegression()
linear_model.fit(X_train_set,y_train_set)

y_pred=linear_model.predict(X_test)

validation_MSE=mean_squared_error(y_pred,y_test)
print(f"Number of training data samples is {Train_set_size} and the Validation MSE is
{Validation_MSE}")
```

P-3: Distance travelled in free fall is given by $d = \frac{g}{2}t^2$ where g is the acceleration of gravity, d denotes the distance, and t denotes the time. You are planning to estimate g using regression. Someone has designed an experiment to measure the distance travelled in free fall and to record corresponding time. He/she has tried with four different values of distances and recorded t^2 for each trial (three trials for each distance) from his/her experiment as listed below:

d (cm)	100	100	100	127	127	127	152	152	152	178	178	178
t^2 (s^2)	0.36	0.38	0.46	0.46	0.49	0.51	0.50	0.53	0.56	0.55	0.58	0.61

1) You are going to design regression model to fit a line through the given data. Write python code to make a scatter plot of the given data. You need to plot t^2 in y-axis and distance *d* in x-axis. That means t^2 is the output (y_p) of your regression model and d is the input ($x_{p,1}$) of your regression model. (**Mark 1**)
2) Calculate the parameters (slope and intercept) that minimize the mean square error (MSE). Print these parameters' values with four digits after decimal point. **Mark 1**
3) From the calculated optimal slope in Q-3, estimate *g* [hint: $d = \frac{g}{2}t^2$]. Compare with the typical value of *g* the acceleration of gravity. **Mark: 3**

P-4: Say you were given with *P*=20 data points. You used polynomial regression to model the relationship between input and output. At first, you used polynomial of order *D=2*. The parameters that minimize mean square error (MSE) for this polynomial regression are w_0=5, w_1=1.5, and w_2=0.03. Then you used polynomial order of *D=3*. The parameters for this polynomial regression are w_0=2, w_1=0.5, w_2=0.01, and w_3=-0.001 that minimize MSE. You are also given 4 data points to validate and compare your models' results in the following table.

Input feature x_1	Output label y_1
4	10
8	20
9	25
14	35

1. Calculate validation MSE for the given data set for both of your regression models (models with *D=2* and *D=3*) **(Mark 2**)
2. Compare your models' validation MSEs and make a conclusion which one of these models (model with *D=2* or model with *D=3*) is more appropriate for your regression problem. (**Mark 2**)

P-5: You have learnt gradient descent algorithm to optimize a function. In this problem you will design hyper parameter which is a learning rate (α). Consider a function $f(w_0) = {w_0}^4 -$

$5{w_0}^2 - 3w_0$. You plan to find the minimum for this function using gradient descent algorithm. In this problem, you will investigate the effect of starting point for w_0 for gradient descent iteration. Assume your initial guess/starting point $w_0 = -2.0$.

1. Calculate w_0 for next two iterations assuming step size $\alpha = 0.1$. Show all steps. (**Mark 1**)
2. Calculate the optimal w_0 using gradient descent for a given learning rate. Try different values of α e.g., 0.2, 0.1, 0.01, 0.001 (**assume initial value of** $w_0 = -2.0$). (**Mark 1**)
   a. Report the optimal value of w_0 for each α (if there is no optimal value is obtained for a given α, you can mention "no optimal value is obtained") and the minimum value of the function for each α using the gradient descent algorithm. (**Mark 1**)
   b. For your answer in a (previous question), did you obtain the **global optimal** value of w_0? If not, mention and show how would you get the global optimal value of w_0? (**Mark 1.5**)