

Lab Assignment 1: Optimization for Machine Learning

Dr. Md Abu Talhamainuddin Ansary

Write python codes of the following problems:

- (1) Price of house (in Rs) with respect to area (sq feet) is give in 2 column excel sheet. Using linear regression, find the price of a house with R -thousand square feet where R is the last digit of your roll no. If last digit of your roll no is 0, choose $R = 2.5$.
- (2) Using 2, 3, 4, 5, 6, 7 degree polynomial fitting curve find the price of a house with R -thousand square feet where R is the last digit of your roll no. If last digit of your roll no is 0, choose $R = 2.5$.
- (3) Price of house based on area and no of bedrooms are given in 3 column excel sheet. Find the best fitting plane $z = a_1x + a_2y + a_3$. Using this find the price of a house with R -thousand square feet and $R + 3$ bedrooms where R is the last digit of your roll no. If last digit of your roll no is 0, choose $R = 2.5$ in house area.
Hint: Use the least square problem $\min_{a \in \mathbb{R}^3} \frac{1}{m} \|Aa - z\|^2$ where $A(i, :) = (x_i, y_i, 1)$.
- (4) Find the best fitting surface $z = a_1x^2 + a_2xy + a_3y^2 + a_4x + a_5y + a_6$ of above data set. Using this find the price of a house with R -thousand square feet and $R + 3$ bedrooms where R is the last digit of your roll no. If last digit of your roll no is 0, choose $R = 2.5$ in house area.
Hint: Use the least square problem $\min_{a \in \mathbb{R}^6} \frac{1}{m} \|Aa - z\|^2$ where $A(i, :) = (x_i^2, x_iy_i, y_i^2, x_i, y_i, 1)$.
- (5) Price of house based on area, bedrooms, and bathrooms are given in four column data. Find the best fitting hyperplane $y = a^T x + \beta$. Using this find the price of a house with R -thousand square feet and $R + 3$ bedrooms where R is the last digit of your roll no and no.of bathrooms is 2 if R is odd and no. of bathrooms is 1 if R is even. If last digit of your roll no is 0, choose $R = 2.5$ in house area.
Hint: Use the least square problem $\min_{x \in \mathbb{R}^4} \frac{1}{m} \|A(x^T, \beta)^T - z\|^2$ where $A(i, :) = (a_i^T, 1)$ where $a_i^T = (area, bedrrom, bathroom)$. Then use $z = a^T x + \beta$.
- (6) From five and six column data find least square solution with 'price' as y .
- (7) Using the census data of Chandigarh predict the population of chandigarh in 2021. (Use exponential fitting).
- (8) **Home work** Define a function *polynomial_fit*(x, y, p) to generate p degree polynomial. Using this justify which degree polynomial gives best approximation among 1 to 20 degree polynomials i.e. which degree polynomial has lowest average loss.

Sample code for linear regression

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dt=pd.read_excel(r'C:\Teaching & Research\Courses\2022-23\Optimization for ML\Lab assignments\'
r'\Housing price and Census datasets\xls\Census data (Chandigarh).xls')
B=dt.values
x,y=B[:,0],B[:,1]
A=np.column_stack((A,np.ones((len(x),1),dtype=float)))
print(A)
beta=np.dot(np.linalg.inv(np.dot(A.T,A)),np.dot(A.T,y.T))
print(beta)
plt.figure(figsize = (min(x)-1,max(x)+1))
plt.plot(x, y, 'b.')
plt.plot(x, beta[0]*x+beta[1], 'r')
plt.xlabel('x')
plt.ylabel('y')
plt.show()
```