 **Centre of Excellence in Artificial Intelligence**

**AI42001:Machine Learning Foundations and Applications**

**Name**-Vaibhav Gupta **Roll Number**-20IE10041

**Date:**30-01-2024 **Assignment -**3

**a. Experiment 1: A table containing the first 10 rows of dataset\_altered.**

|  | **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **MEDV** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 0.00632 | 18.0 | 2.31 | 0.0 | 0.538 | 6.575 | 65.2 | 4.0900 | 1.0 | 296.0 | 15.3 | 24.0 |
| **1** | 0.02731 | 0.0 | 7.07 | 0.0 | 0.469 | 6.421 | 78.9 | 4.9671 | 2.0 | 242.0 | 17.8 | 21.6 |
| **2** | 0.02729 | 0.0 | 7.07 | 0.0 | 0.469 | 7.185 | 61.1 | 4.9671 | 2.0 | 242.0 | 17.8 | 34.7 |
| **3** | 0.03237 | 0.0 | 2.18 | 0.0 | 0.458 | 6.998 | 45.8 | 6.0622 | 3.0 | 222.0 | 18.7 | 33.4 |
| **4** | 0.06905 | 0.0 | 2.18 | 0.0 | 0.458 | 7.147 | 54.2 | 6.0622 | 3.0 | 222.0 | 18.7 | 36.2 |
| **5** | 0.02985 | 0.0 | 2.18 | 0.0 | 0.458 | 6.430 | 58.7 | 6.0622 | 3.0 | 222.0 | 18.7 | 28.7 |
| **7** | 0.14455 | 12.5 | 7.87 | 0.0 | 0.524 | 6.172 | 96.1 | 5.9505 | 5.0 | 311.0 | 15.2 | 27.1 |
| **8** | 0.21124 | 12.5 | 7.87 | 0.0 | 0.524 | 5.631 | 100.0 | 6.0821 | 5.0 | 311.0 | 15.2 | 16.5 |
| **10** | 0.22489 | 12.5 | 7.87 | 0.0 | 0.524 | 6.377 | 94.3 | 6.3467 | 5.0 | 311.0 | 15.2 | 15.0 |
| **11** | 0.11747 | 12.5 | 7.87 | 0.0 | 0.524 | 6.009 | 82.9 | 6.2267 | 5.0 | 311.0 | 15.2 | 18.9 |

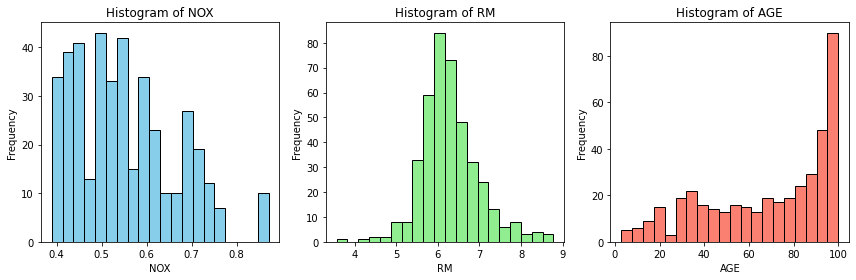
**b. Experiment 2: Histograms of “NOX”, “RM” and “AGE” for dataset\_altered; table containing correlation coefficients; correlation matrix heatmap. State what all you can infer from the correlation matrix.**

Table containing correlation coefficients

CRIM ZN INDUS CHAS NOX RM AGE \

CRIM 1.000000 -0.187914 0.399125 -0.053812 0.426892 -0.231273 0.347881

ZN -0.187914 1.000000 -0.517127 -0.017860 -0.508431 0.334985 -0.556569

INDUS 0.399125 -0.517127 1.000000 0.042823 0.764716 -0.406520 0.639033

CHAS -0.053812 -0.017860 0.042823 1.000000 0.066370 0.096277 0.059016

NOX 0.426892 -0.508431 0.764716 0.066370 1.000000 -0.317127 0.734957

RM -0.231273 0.334985 -0.406520 0.096277 -0.317127 1.000000 -0.253154

AGE 0.347881 -0.556569 0.639033 0.059016 0.734957 -0.253154 1.000000

DIS -0.369388 0.643411 -0.698102 -0.086243 -0.767751 0.221341 -0.747306

RAD 0.613026 -0.298081 0.600663 0.007043 0.639015 -0.236786 0.452883

TAX 0.566329 -0.303079 0.738828 -0.033658 0.687973 -0.322061 0.511233

PTRATIO 0.268360 -0.420645 0.397222 -0.109588 0.215724 -0.386630 0.263088

MEDV -0.398374 0.395418 -0.511117 0.174342 -0.459843 0.728768 -0.411209

DIS RAD TAX PTRATIO MEDV

CRIM -0.369388 0.613026 0.566329 0.268360 -0.398374

ZN 0.643411 -0.298081 -0.303079 -0.420645 0.395418

INDUS -0.698102 0.600663 0.738828 0.397222 -0.511117

CHAS -0.086243 0.007043 -0.033658 -0.109588 0.174342

NOX -0.767751 0.639015 0.687973 0.215724 -0.459843

RM 0.221341 -0.236786 -0.322061 -0.386630 0.728768

AGE -0.747306 0.452883 0.511233 0.263088 -0.411209

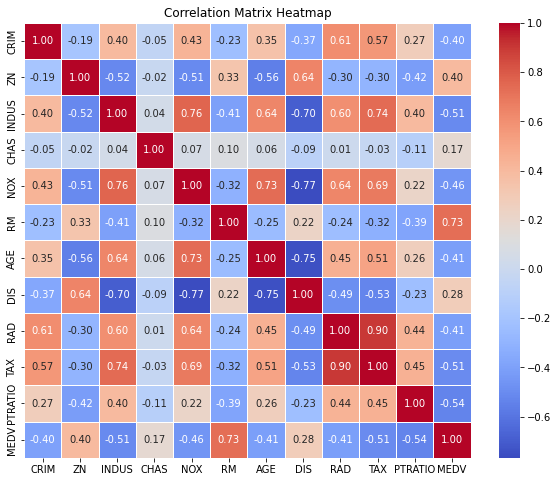
DIS 1.000000 -0.485381 -0.532896 -0.229773 0.279111

RAD -0.485381 1.000000 0.902564 0.440476 -0.411004

TAX -0.532896 0.902564 1.000000 0.446342 -0.505214

PTRATIO -0.229773 0.440476 0.446342 1.000000 -0.537074

MEDV 0.279111 -0.411004 -0.505214 -0.537074 1.000000

Correlation heatmap:

We can infer from the heatmap that MEDV (median value of the house) is mostly correlated with RM( average number of rooms) with highest correlation 0.73 and negative correlated with PTRATIO(pupil-teacher ratio by town),INDUS,TAX i.e, MEDV decreases if PTRATIO or INDUS or TAX increases and vice versa

MEDV depends least on CHAS(Charles River dummy variable) with correlation 0.17.

**c. Experiment 3: Print the shape of individual data matrices.**

dataset\_altered\_features is 412\*11

dataset\_altered\_target is 412\*1

**d. Experiment 4: Values for coefficients and intercept; RMSE value of predicted**

**data with the testing data.**

[22.52601114 -1.49672144 0.87988611 -0.06764907 0.80887556 -2.56486945

4.29763482 -1.39058959 -3.09460772 2.07264836 -2.03014342 -2.13404532]

These are the coefficient values for the closed form solution of linear regression where 22.52(first term) is the bias and the rest terms are coefficients for the other 11 columns

| **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

respectively .

Root Mean Squared Error: 3.784652957464752

**e. Experiment 5: A bar plot of RMSEs vs learning rate. State the optimal learning**

**rate, and corresponding values for coefficients and intercept.**

Root Mean Squared Error for learning rate 0.001 = 4.830276105041903

Root Mean Squared Error for learning rate 0.01 = 3.8328100192668404

Root Mean Squared Error for learning rate 0.1 = 3.78465376845852

**Optimal Learning Rate:** 0.1

Corresponding RMSE: 3.78465376845852

**Optimal Weights:** [-1.49672082 0.87988475 -0.06765266 0.80887601 -2.56486834 4.29763529

-1.39059007 -3.09460759 2.07264024 -2.03013396 -2.13404512]

for the 11 columns

| **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

respectively .

**Optimal Bias:** 22.526011234276915

Bar plot of RMSE vs learning rates 0.001,0.01,0.1 after 1000 iterations of gradient descent

