**CST8390 - Lab 5 Regression**

**Answer Document**

1 a. Number of instances: 513

b. Number of attributes: 13

c. Attribute information:

|  |  |
| --- | --- |
| Attribute | Description |
| CRIM | Per capita crime rate by town |
| ZN | proportion of residential land zoned for lots over 25,000 sq.ft. |
| INDUS | proportion of non-retail business acres per town |
| CHAS | Charles River dummy (variable = 1 if tract bounds river; 0 otherwise) |
| NOX | nitric oxides concentration (parts per 10 million) |
| RM | average number of rooms per dwelling |
| AGE | proportion of owner-occupied units built prior to 1940 |
| DIS | weighted distances to five Boston employment centres |
| RAD | index of accessibility to radial highways |
| TAX | full-value property-tax rate per $10,000 |
| PTRATIO | pupil-teacher ratio by town |
| LSTAT | % lower status of the population |
| MEDV | Median value of owner-occupied homes in $1000's |

2 a. 21.1 b. 8.704 c. 0.332

4 a. simple linear regression model

b. average number of room per dwelling (RM) and if tract bound to river (Chas = 1)

c. NOX nitric oxides concentration (parts per 10 million), and if tract is not bound to river chas = 0

5.

