ISCO630E

*Analysis Report on **Assignment 2***

Group Members :

1. *Sashank Mishra - IIT2016515*
2. *Shreyansh Dwivedi - IWM2016501*
3. *Saurabh Kumar - IWM2016502*
4. *Shreyansh Chaudhary - IIT2016068*
5. *Pradeep Gangwar - IHM2016501*

QUESTION 1

Dataset : Housing Price Prediction (in CSV format)
546 rows, 12 columns

price	lotsize	bedrooms	bathrms	stories	driveway	recroom	fullbase	gashw	airco	garagepl	prefarea
42000.0	5850	3	1	2	yes	no	yes	no	no	1	no
38500.0	4000	2	1	1	yes	no	no	no	no	0	no
49500.0	3060	3	1	1	yes	no	no	no	no	0	no
60500.0	6650	3	1	2	yes	yes	no	no	no	0	no
61000.0	6360	2	1	1	yes	no	no	no	no	0	no

Data Preprocessing :

- All the columns containing the option as ‘yes’ and ‘no’ are mapped to 0 and 1, respectively.
  The final data representation:

price	lotsize	bedrooms	bathrms	stories	driveway	recroom	fullbase	gashw	airco	garagepl	prefarea
42000.0	5850	3	1	2	1	0	1	0	0	1	0
38500.0	4000	2	1	1	1	0	0	0	0	0	0
49500.0	3060	3	1	1	1	0	0	0	0	0	0
60500.0	6650	3	1	2	1	1	0	0	0	0	0
61000.0	6360	2	1	1	1	0	0	0	0	0	0

- Data splitted into features and prediction(X and Y, respectively)
  Prediction variable here is the ‘price’
- Data is normalized, and a Bias term added to feature set.

Analysis :

Normal Equation with Regularization:

Lambda	Final Cost
1	0.16376182676447748
100	0.24522819472882595
500	0.4111756460232463
1000	0.5330343556160958
1500	0.6406573247629825

Gradient Descent with Regularization

No. of iterations	Learning Rate	Lambda	Final Cost
1000	0.001	10	0.6356541779944964
10000	0.001	10	0.4620079008107054
10000	0.001	500	385.16088606256363
10000	0.01	1	0.36317156005140366

QUESTION 2

Dataset : Housing Price Prediction (in CSV format)
546 rows, 12 columns

price	lotsize	bedrooms	bathrms	stories	driveway	recroom	fullbase	gashw	airco	garagepl	prefarea
42000.0	5850	3	1	2	yes	no	yes	no	no	1	no
38500.0	4000	2	1	1	yes	no	no	no	no	0	no
49500.0	3060	3	1	1	yes	no	no	no	no	0	no
60500.0	6650	3	1	2	yes	yes	no	no	no	0	no
61000.0	6360	2	1	1	yes	no	no	no	no	0	no

Data Preprocessing :

- All the columns containing the option as ‘yes’ and ‘no’ are mapped to 0 and 1, respectively.
  The final data representation:

price	lotsize	bedrooms	bathrms	stories	driveway	recroom	fullbase	gashw	airco	garagepl	prefarea
42000.0	5850	3	1	2	1	0	1	0	0	1	0
38500.0	4000	2	1	1	1	0	0	0	0	0	0
49500.0	3060	3	1	1	1	0	0	0	0	0	0
60500.0	6650	3	1	2	1	1	0	0	0	0	0
61000.0	6360	2	1	1	1	0	0	0	0	0	0

- Data splitted into features and prediction(X and Y, respectively)
  Prediction variable here is the ‘price’
- Data is normalized, and a Bias term added to feature set.

Analysis :

Value of tau	Error
1	0.07539028
10	0.3233095
100	0.32684042
1000	0.32687602
0.1	0.01345146
0.01	0.00509697
0.001	0.00386234

It is clear that as we decrease the value of tau, the error decreases, and so we get a better fit to the data. But, then we also get into the risk of overfitting.