Machine Learning

Dr Changjiang He, Dr Kuo-Ming Chao
Computer Science| School of Art
University of Roehampton

Lesson 2.2

Linear Regression

Contents

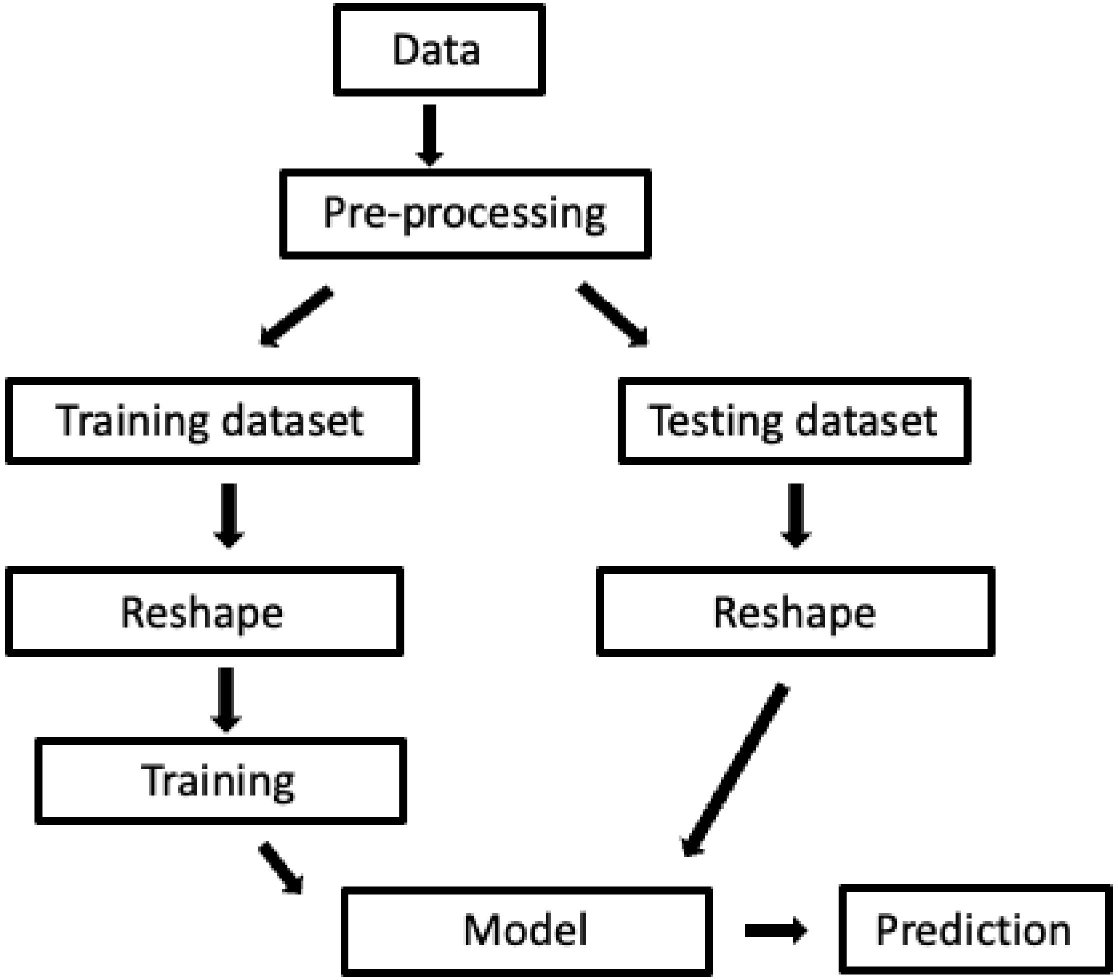
- General process of supervised learning
- What is linear regression?
- How to calculate mean squared error?
- What method can be used to estimate parameters of linear regression model?
- Other regression methods

Basic Framework of Regression

Most regression models follow similar process.

For example:

- Linear Regression
- Lasso
- Neural Network
- …



Linear Regression

- A **paramaterized** prediction model is initialised with a set of random parameters and an error function is used to judge how well this initial model performs when making predictions for instances in a training dataset.
- Based on the value of the error function the parameters are iteratively adjusted to create a more and more accurate model.

Linear Regression is a simple parameterised predication model

Linear Regression

University of Roehampton London

In the simplest case, the regression model allows for a linear relationship.

$$y = m \times x + b$$

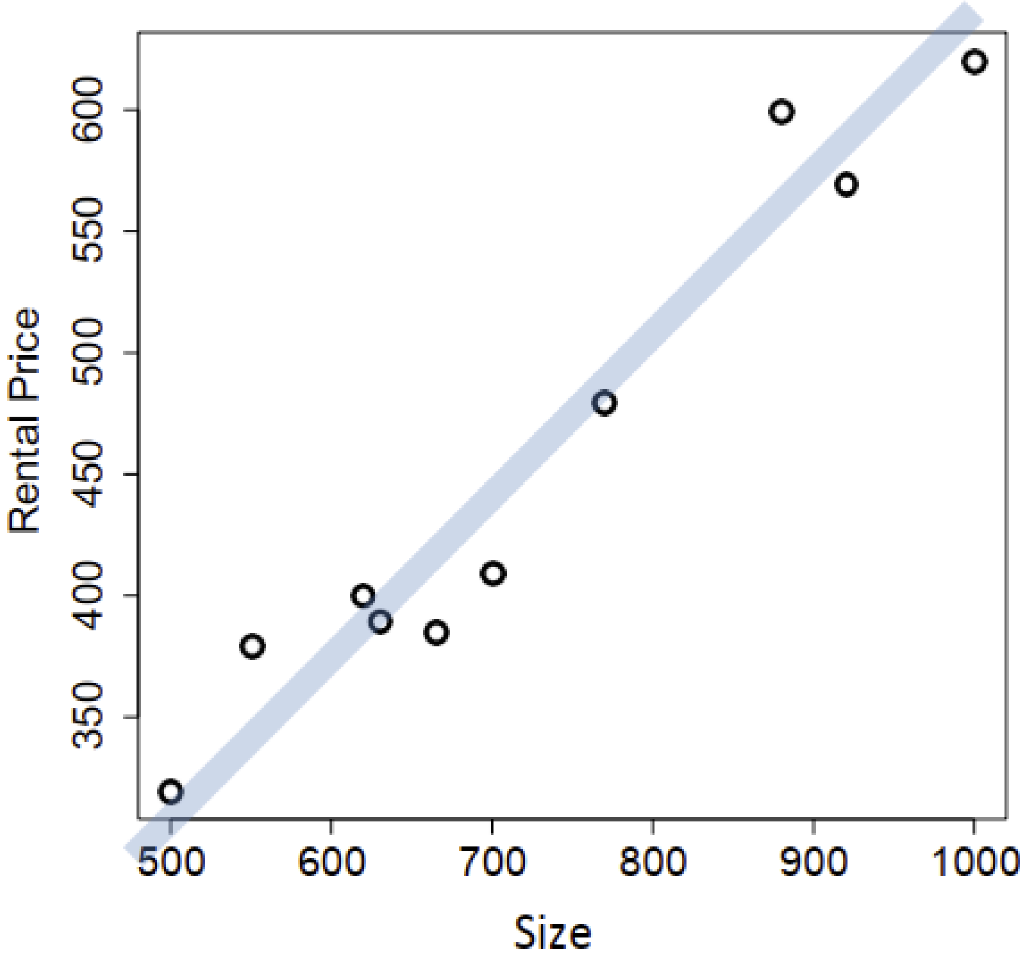
y is the forecast variable – rental price

x is the predictor - size

m, b are parameters;

ID	Size	Rental Price
1	500	320
2	550	380
3	620	400
4	630	390
5	665	385
6	700	410
7	770	480
8	880	600
9	920	570
10	1,000	620

Linear Regression



A scatter plot of the SIZE and RENTAL PRICE features from the office rentals dataset.

ID	SIZE	RENTAL PRICE
1	500	320
2	550	380
3	620	400
4	630	390
5	665	385
6	700	410
7	770	480
8	880	600
9	920	570
10	1,000	620

The model

$$y = m \times x + b$$

uses a straight line to fit the data

How to estimate m and b?

Linear Regression

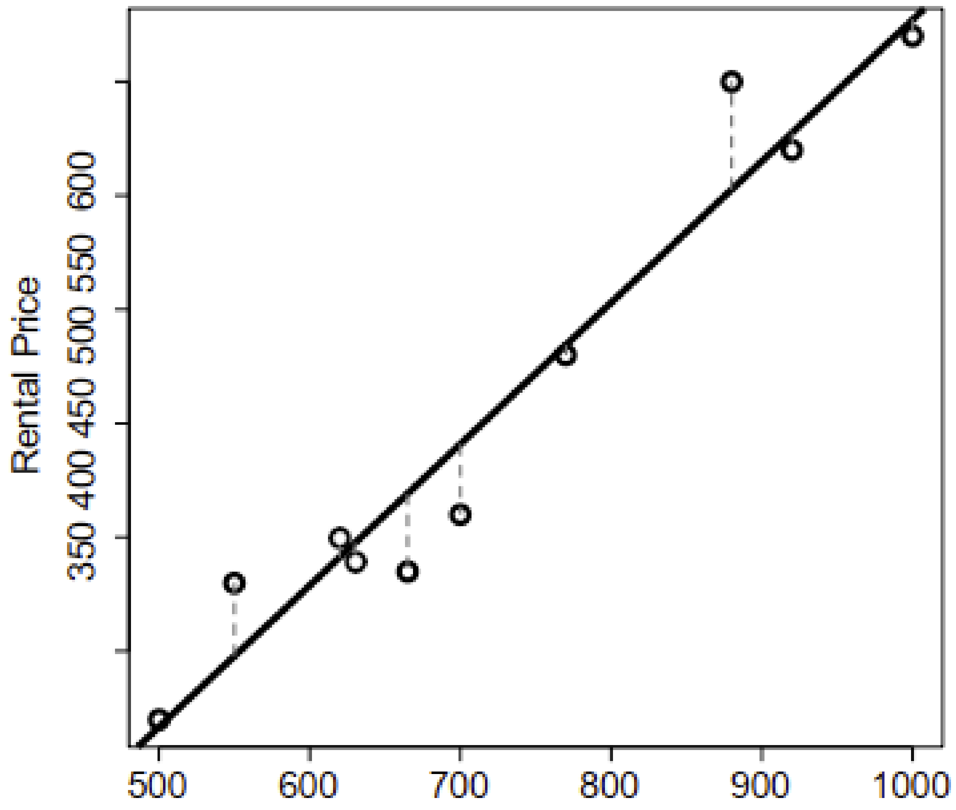
First lets think of a question:

Assume that we have calculated the parameters and the model is:

$$\text{RENTAL PRICE} = 6.47 + 0.62 \times \text{SIZE}$$

How do we know if the model is accurate or not?

We can calculate the errors!



A candidate prediction model (with w[0] = 6.47 and w[1] = 0.62) and the resulting errors.

Linear Regression: Errors

Error is the difference between true values and predicted values $Y_i - \hat{Y}_i$

ID	Size	Rental Price
1	500	320
2	550	380
3	620	400
4	630	390
5	665	385
6	700	410
7	770	480
8	880	600
9	920	570
10	1,000	620

RENTAL PRICE = 6.47 + 0.62 × SIZE

ID	Rental Price	Model Prediction	Error	Squared Error
1	320	316.79	3.21	10.32
2	380	347.82	32.18	1,035.62
3	400	391.26	8.74	76.32
4	390	397.47	-7.47	55.80
5	385	419.19	-34.19	1,169.13
6	410	440.91	-30.91	955.73
7	480	484.36	-4.36	19.01
8	600	552.63	47.37	2,243.90
9	570	577.46	-7.46	55.59
10	620	627.11	-7.11	50.51

Linear Regression: Errors

Squared Error is the square of the error $(Y_i - \hat{Y}_i)^2$

ID	Size	Rental Price
1	500	320
2	550	380
3	620	400
4	630	390
5	665	385
6	700	410
7	770	480
8	880	600
9	920	570
10	1,000	620

RENTAL PRICE = 6.47 + 0.62 × SIZE

ID	Rental Price	Model Prediction	Error	Squared Error
1	320	316.79	3.21	10.32
2	380	347.82	32.18	1,035.62
3	400	391.26	8.74	76.32
4	390	397.47	-7.47	55.80
5	385	419.19	-34.19	1,169.13
6	410	440.91	-30.91	955.73
7	480	484.36	-4.36	19.01
8	600	552.63	47.37	2,243.90
9	570	577.46	-7.46	55.59
10	620	627.11	-7.11	50.51

Linear Regression: Errors

Then, we can calculate sum of squared errors $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$

ID	Size	Rental Price
1	500	320
2	550	380
3	620	400
4	630	390
5	665	385
6	700	410
7	770	480
8	880	600
9	920	570
10	1,000	620

Rental Price = 6.47 + 0.62 × Size

ID	Rental Price	Model Prediction	Error	Squared Error
1	320	316.79	3.21	10.32
2	380	347.82	32.18	1,035.62
3	400	391.26	8.74	76.32
4	390	397.47	-7.47	55.80
5	385	419.19	-34.19	1,169.13
6	410	440.91	-30.91	955.73
7	480	484.36	-4.36	19.01
8	600	552.63	47.37	2,243.90
9	570	577.46	-7.46	55.59
10	620	627.11	-7.11	50.51
			Sum	5,671.64

Linear Regression: Errors

Finally, we calculate mean squared errors (MSE) $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$

ID	Size	Rental Price
1	500	320
2	550	380
3	620	400
4	630	390
5	665	385
6	700	410
7	770	480
8	880	600
9	920	570
10	1,000	620

Rental Price = 6.47 + 0.62 $\times$ Size

ID	Rental Price	Model Prediction	Error	Squared Error
1	320	316.79	3.21	10.32
2	380	347.82	32.18	1,035.62
3	400	391.26	8.74	76.32
4	390	397.47	-7.47	55.80
5	385	419.19	-34.19	1,169.13
6	410	440.91	-30.91	955.73
7	480	484.36	-4.36	19.01
8	600	552.63	47.37	2,243.90
9	570	577.46	-7.46	55.59
10	620	627.11	-7.11	50.51
			Sum	5,671.64

MSE

$= \frac{5671.64}{10}$

$= 567.164$

Linear Regression

Now we can answer the question of how to estimate the parameter of the linear regression model.

We can use a mathematical method, called least square estimation, to estimate the parameters by minimizing the mean squared errors MSE.

That is, we want to minimise $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$

In seminar sheet we will learn how to use least square estimation to build a linear regression model