Getting Started with Machine Learning

Raunak J

In last tutorial

We learned R-Squared method.

Its mathematical example.

Python Implementation too.

What if want to find the Goodness of Fit without the prediction values ?

That is just by using X and Y training data.

No prediction mechanism at all.

Is there a way ?

Definitely there is…

$$Formula\ we\ have\ for\ R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2}$$

$$But\ if\ you\ look\ at\ R^2, it\ is\ also\ denoted\ as\ r^2$$

$$So\ with\ what\ can\ we\ denote\ the\ r\ ?$$

$$If\ you\ recall\ it\ is\ also\ known\ as\ Karl\ Pearson's\ Coefficient\ of\ Correlation.$$

So the Karl Pearson's Co-efficient of Correlation is denoted by :

$$r = \frac{\sum(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2}}$$

$$r = \frac{Cov(x, y)}{s_x * s_y} = \frac{\frac{\sum(x_i - \bar{x}).(y_i - \bar{y})}{N - 1}}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{N - 1}} * \sqrt{\frac{\sum(y_i - \bar{y})^2}{N - 1}}}$$

*So the Coefficient of Determination can be given by r*r*

The intuition can be given better with an example.

Height of the person in cms (x)	**Weight of the person in kgs (y)**
160	72
171	76
182	77
180	83
154	76

The mean is given for following data

Height of the person in cms (x)	Weight of the person in kgs (y)
160	72
171	76
182	77
180	83
154	76

$$\bar{x} = \frac{\sum x_i}{N}$$

$$\bar{y} = \frac{\sum y_i}{N}$$

$$\bar{x} = \frac{160 + 171 + 182 + 180 + 154}{5}$$

$$\bar{y} = \frac{72 + 76 + 77 + 83 + 76}{5}$$

$$\bar{x} = 169.4$$

$$\bar{y} = 76.8$$

X	Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}).(y_i - \bar{y})$
160	72	-9.4	-4.8	88.36	23.04	45.12
171	76	1.6	-0.8	2.56	0.64	-1.28
182	77	12.6	0.2	158.76	0.04	2.52
180	83	10.6	6.2	112.36	38.44	65.72
154	76	-15.4	-0.8	237.16	0.64	12.32
$\Sigma 847$	$\Sigma 384$	$\Sigma 0$	$\Sigma 0$	$\Sigma 599.2$	$\Sigma 62.8$	$\Sigma 124.4$

$$r = \frac{\Sigma(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2.\Sigma(y_i - \bar{y})^2}} = \frac{124.4}{\sqrt{(599.2) * (62.8)}} = 0.6413$$

X	Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}).(y_i - \bar{y})$
160	72	-9.4	-4.8	88.36	23.04	45.12
171	76	1.6	-0.8	2.56	0.64	-1.28
182	77	12.6	0.2	158.76	0.04	2.52
180	83	10.6	6.2	112.36	38.44	65.72
154	76	-15.4	-0.8	237.16	0.64	12.32
$\Sigma 847$	$\Sigma 384$	$\Sigma 0$	$\Sigma 0$	$\Sigma 599.2$	$\Sigma 62.8$	$\Sigma 124.4$

$$Cov(x,y) = \frac{\sum(x_i - \bar{x}).(y_i - \bar{y})}{N-1} = \frac{124.4}{4} = 31.1$$

$$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{599.2}{4}} = 12.24$$

$$s_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{N-1}} = \sqrt{\frac{62.8}{4}} = 3.962$$

X	Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}).(y_i - \bar{y})$
160	72	-9.4	-4.8	88.36	23.04	45.12
171	76	1.6	-0.8	2.56	0.64	-1.28
182	77	12.6	0.2	158.76	0.04	2.52
180	83	10.6	6.2	112.36	38.44	65.72
154	76	-15.4	-0.8	237.16	0.64	12.32
$\Sigma 847$	$\Sigma 384$	$\Sigma 0$	$\Sigma 0$	$\Sigma 599.2$	$\Sigma 62.8$	$\Sigma 124.4$

$$r = \frac{Cov(x,y)}{s_x * s_y} = \frac{\frac{\sum(x_i - \bar{x}).(y_i - \bar{y})}{N-1}}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{N-1}} * \sqrt{\frac{\sum(y_i - \bar{y})^2}{N-1}}} = \frac{31.1}{12.24 * 3.962} = 0.6413$$

X	Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}).(y_i - \bar{y})$
160	72	-9.4	-4.8	88.36	23.04	45.12
171	76	1.6	-0.8	2.56	0.64	-1.28
182	77	12.6	0.2	158.76	0.04	2.52
180	83	10.6	6.2	112.36	38.44	65.72
154	76	-15.4	-0.8	237.16	0.64	12.32
$\Sigma 847$	$\Sigma 384$	$\Sigma 0$	$\Sigma 0$	$\Sigma 599.2$	$\Sigma 62.8$	$\Sigma 124.4$

$$Final\ step\ is\ finding\ r^2 = .6413^2 = .41127$$