

حصصی بیت رستم آمار احتمال مهندسی دکتر سرفی زاربی - دانشکده سرفی

رگرسیون خطی (Linear Regression)

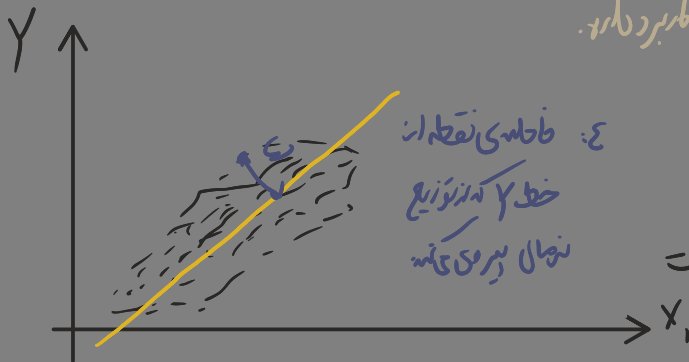
یک سری داده داریم که ویژگی های زیادی دارند مثل جدول در به دو

	x_1	x_2	x_3	...	x_n	y
1						
2						
...						
n						

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

به این کار یک ویژگی بر حسب بقیه ویژگی ها.

نه یادگیری ماشین و پس بینی کردن کاربرد دارد.



به نسبت به معلوم داده و یادگیری ماشین کاربرد دارد

هدف رسم یک خط y که بیشترین برای مقادیر x_1, x_2, \dots, x_n است

$$y = \beta_0 + \beta_1 x + \epsilon$$

$\epsilon \sim N(0,1)$

دلیل سازیم کردن احتمال وقوع این داده ها برای خطی که رسم کردیم \Leftarrow Likelihood

$$L(y_i | \beta_0, \beta_1, x_i, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}$$

$$\Rightarrow L(\text{Data} | \theta) = \text{MLE}_{\beta_0, \beta_1} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}$$

از قبل یاد گرفته ایم که برای هر دو پارامتر β_0 و β_1 می توانیم به صورت جداگانه به دست آوریم.

$$\beta_0 = \bar{y} - \beta_1 \bar{x}, \quad \beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\sigma^2 = \frac{1}{n} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\Rightarrow \beta_0 = \bar{y} - \beta_1 \bar{x} \quad , \quad \beta_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \quad , \quad \sigma^2 = \frac{1}{n} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$