Learning from data: Predictive Modelling

Linear Regression

Input features, targets.

Quantitative variables, Qualitative (Categorical) Variables

Predictive Modelling f: Input features $\longrightarrow$ targets.

1. We want to predict if a patient is likely to require Caesarian delivery.

What information will be useful to predict this?

Input features: Patient Age, Patient history for example if this is the first ~~time~~ pregnancy? ~~If the patient has undergone~~

Target: requires caesarian delivery vs. does not require caesarian delivery

The target here is a categorical variable. {Yes, No}

The input feature "Patient Age" is a quantitative $age \in \mathbb{N}$.

2. We want to predict the value of a stock in the stock market.

Input features: Past history of stock, Current financial situation of company, future growth prospective.

Stock Prediction: This is an example of regression problem

$f: X \rightarrow y \quad y \in \mathbb{R}$

$\mathbb{R}$: Real number.

Catsaraian delivery : This is an example of classification problem.

$f: x \rightarrow y \quad y \in \{0, 1, 2, \dots k\}$.

where the target/label is a category.

Linear Regression

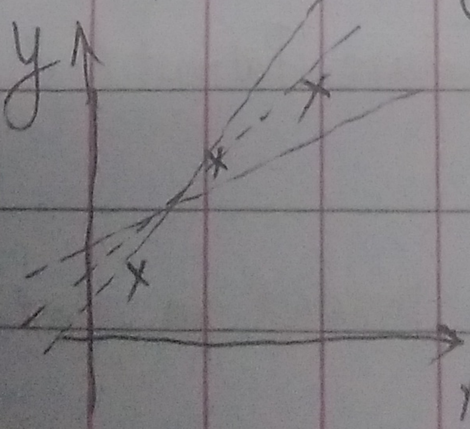
$f: X \rightarrow y \quad y \in \mathbb{R}$.

f: is linear.

Lets take a simple example of learning to predict from data where input and target are scalar variables.

We will fit a line of the form $y = \theta_0 + \theta_1 x$.

x	y
1	4.8
3	11.3
5	17.2
6	?

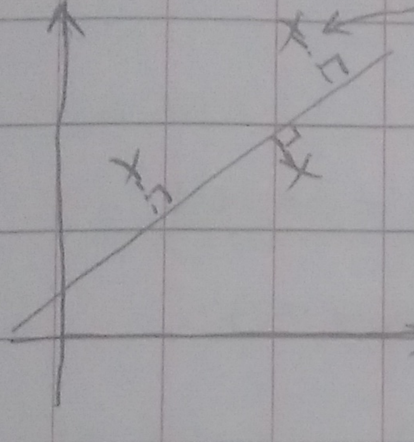


All these dotted lines look like good enough fit for the data.

We want to pick a unique solution.

What would be a good criteria of choosing among all these probable lines?

Solution: One reasonable solution is to pick a line s.t. Mean square error (MSE) is smallest.



$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2$$

(also called L2 norm) of the error.

Other reasonable solution

Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}|$$

(also called L1-norm)

To obtain the parameters (θ_0, θ_1) of the line that satisfies the least square error, use the following formula

$$\theta_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2} \qquad \theta_0 = \bar{y} - \theta_1 \cdot \bar{x} \quad \text{—— (1)}$$

TODO: derivation of the equation (1).

Applying eqn (1), we have

x	y	$x \cdot y$	x^2
1	4.8	4.8	1
3	11.3	33.9	9
5	17.2	86	25
		35	

Average:

$\bar{x} = 3, \bar{y} = 11.1, \overline{xy} = 41.57, \overline{x^2} = 11.67.$

$$\theta_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{41.57 - 3 \times 11.1}{11.67 - 9}$$

$$\doteq 3.1$$

$$\theta_0 = \bar{y} - \theta_0 \cdot \bar{x} = 11.1 - 3.1 \times 3.$$

$$\doteq 1.8.$$

Thus, the best fit line for the MSE criteria is

$$\boxed{y = f(x) = 1.8 + 3.1x}$$

Derivation of Equation 1 -

* Where did this magical equation ① come from?

Cost function

Given a dataset $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, find a functional relationship of the form

$$f(x;\theta) = f_\theta(x) = \theta_0 + \theta_1 x$$

such that $f_\theta(x)$ minimizes mean squared error.

i.e cost function $\min_\theta L(f_\theta(x), D)$

Solution Use the idea that $\frac{\partial}{\partial \theta} L(f_\theta(x), D) = 0$ at the point where $\theta = \theta^*$ optimal.

$$\frac{\partial}{\partial \theta} L(f_\theta(x), D) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} y_i - (\theta_0 + \theta_1 x_i)^2$$

Differentiating the cost function,

$$\frac{\partial L}{\partial \theta_0} = \frac{1}{N} \sum_{i=1}^{N} (-2y_i + 2\theta_0 + 2\theta_1 x_i) \quad -②$$

$$\frac{\partial L}{\partial \theta_1} = \frac{1}{N} \sum_{i=1}^{N} (-2x_i y_i + 2\theta_0 x_i + 2\theta_1 x_i^2) \quad -③$$

Equating $\frac{\partial L}{\partial \theta_0} = 0$

we have, $\frac{1}{N} \sum_{i=1}^{N} (-y_i + \theta_0 + \theta_1 x_i) = 0$

$\Rightarrow -\bar{y} + \theta_0 + \theta_1 \bar{x} = 0$

$\Rightarrow \theta_0 + \theta_1 \bar{x} - \bar{y} = 0 \quad -(4)$

Equating $\frac{\partial L}{\partial \theta_1} = 0$,

we have, $\frac{1}{N} \sum_{i=1}^{N} (-x_i y_i + \theta_0 x_i + \theta_1 x_i^2) = 0$.

$\Rightarrow -\overline{xy} + \theta_0 \bar{x} + \theta_1 \overline{x^2} = 0$

$\Rightarrow \theta_0 \bar{x} + \theta_1 \overline{x^2} - \overline{xy} = 0 \quad -(5)$

Equating eqn (4) and (5),

$\theta_0 + \theta_1 \bar{x} - \bar{y} = 0$

$\theta_0 \bar{x} + \theta_1 \overline{x_2} - \overline{xy} = 0$

$\Rightarrow \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \overline{xy} \end{bmatrix}$

$$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{bmatrix}^{-1} \begin{bmatrix} \bar{y} \\ \overline{xy} \end{bmatrix}$$

$$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \frac{1}{\overline{x^2} - (\bar{x})^2} \begin{bmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} \bar{y} \\ \overline{xy} \end{bmatrix}$$

$$= \frac{1}{\overline{x^2} - (\bar{x})^2} \begin{bmatrix} \overline{x^2} \cdot \bar{y} - \bar{x} \cdot \overline{xy} \\ -\bar{x} \cdot \bar{y} + \overline{xy} \end{bmatrix}$$

Hence, $\boxed{\theta_1 = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}}$

$\theta_0 = \dfrac{\overline{x^2} \cdot \bar{y} - \bar{x} \cdot \overline{xy}}{\overline{x^2} - (\bar{x})^2}$. This can be simplified further -

$$= \frac{\overline{x^2} \cdot \bar{y} - (\bar{x})^2 \cdot \bar{y} + (\bar{x})^2 \cdot \bar{y} - \bar{x} \cdot \overline{xy}}{\overline{x^2} - (\bar{x})^2}$$

$$= \bar{y} \cdot \frac{\overline{x^2} - (\bar{x})^2}{x^2 - (\bar{x})^2} - \bar{x} \cdot \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}$$

$$\boxed{\theta_0 = \bar{y} - \theta_1 \cdot \bar{x}}$$