Linear Regression

Regressions

Consider the data set $\{(x_i, y_i)\}_{i=1}^{m}$ where y_i are the target output given the inputs x_i .

Let $\hat{y}_i$ be the estimate of y_i , i.e.

$$y_i = \hat{y}_i + \varepsilon_i$$

where ε_i represents the residue for $i = 1, 2, \ldots, m$.

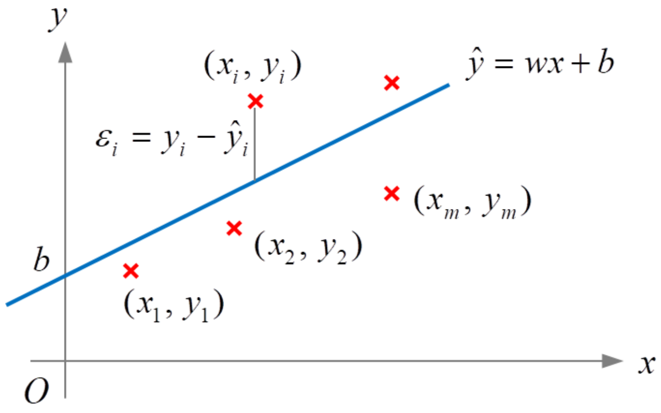
Averaging yields

$$\bar{y} = \bar{\hat{y}} + \bar{\varepsilon}$$

Given the condition that $\bar{\varepsilon} = 0$, we may have

$$\bar{y} = \bar{\hat{y}} \ .$$

Linear Regression



A *regression* (回歸) is a statistical model that relates a dependent variable to the independent variable(s).

A statistical model is called *linear regression* (線性回歸) if

$$\hat{y}_i = wx_i + b$$

where w is called the *weight*, and b is called the *bias*, for $i = 1, 2, \ldots, m$.

Note that $\hat{y}_i = wx_i + b = \underline{x}_i \cdot \underline{w}$ where $\underline{x}_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix}$ and $\underline{w} = \begin{pmatrix} b \\ w \end{pmatrix}$.

Linear Regression

Linear regression tries to find (b, w) that minimizes the *residual sum of squares*,

$$RSS = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \ .$$

$$\begin{aligned}
\because \quad RSS &= \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \\
&= \sum_{i=1}^{m} (y_i - wx_i - b)^2 \\
&= \sum_{i=1}^{m} [(y_i - wx_i)^2 - 2b(y_i - wx_i) + b^2] \\
&= \sum_{i=1}^{m} (y_i - wx_i)^2 - 2b\sum_{i=1}^{m} (y_i - wx_i) + mb^2 \\
&= \sum_{i=1}^{m} (y_i - wx_i)^2 - 2mb(\overline{y} - w\overline{x}) + mb^2 \qquad \left(\because \overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i \ \ \& \ \ \overline{y} = \frac{1}{m}\sum_{i=1}^{m} y_i \right) \\
&= \sum_{i=1}^{m} (y_i - wx_i)^2 + m[b^2 - 2b(\overline{y} - w\overline{x})] \\
&= \sum_{i=1}^{m} (y_i - wx_i)^2 + m\left\{[b - (\overline{y} - w\overline{x})]^2 - (\overline{y} - w\overline{x})^2\right\}
\end{aligned}$$

$\Rightarrow \quad RSS$ is minimized to $RSS = \sum_{i=1}^{m} (y_i - wx_i)^2 - m(\overline{y} - w\overline{x})^2$ if $b = \overline{y} - w\overline{x}$.

Linear Regression

Hence

$$
\begin{aligned}
RSS\big|_{b=\bar{y}-w\bar{x}} &= \sum_{i=1}^{m}(y_i-\hat{y}_i)^2 \\
&= \sum_{i=1}^{m}(y_i-\bar{y}-w(x_i-\bar{x}))^2 \\
&= \sum_{i=1}^{m}[(y_i-\bar{y})^2-2w(x_i-\bar{x})(y_i-\bar{y})+w^2(x_i-\bar{x})^2]^2 \\
&= \sum_{i=1}^{m}(y_i-\bar{y})^2-2w\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})+w^2\sum_{i=1}^{m}(x_i-\bar{x})^2 \\
&= \sum_{i=1}^{m}(y_i-\bar{y})^2+\left[\sum_{i=1}^{m}(x_i-\bar{x})^2\right]\left[w^2-2w\frac{\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{m}(x_i-\bar{x})^2}\right] \\
&= \sum_{i=1}^{m}(y_i-\bar{y})^2+\left[\sum_{i=1}^{m}(x_i-\bar{x})^2\right]\left\{\left[w-\frac{\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{m}(x_i-\bar{x})^2}\right]^2-\left[\frac{\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{m}(x_i-\bar{x})^2}\right]^2\right\}
\end{aligned}
$$

$\Rightarrow$ RSS is further minimized to $RSS=\sum_{i=1}^{m}(y_i-\bar{y})^2-w^2\sum_{i=1}^{m}(x_i-\bar{x})^2$ if $w=\dfrac{\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{m}(x_i-\bar{x})^2}$.

Linear Regression

Hence, $RSS(b, w)$ is the minimum if the following arguments are taken:

$$\begin{cases} b = \overline{y} - w\overline{x} \ , \\ w = \dfrac{\text{cov}(X, Y)}{\text{cov}(X, X)} \ . \end{cases}$$

where $\text{cov}(X, Y)$ is the covariance between X and Y .

Linear Regression

$$\because \quad \hat{y}_i = wx_i + b$$

$$\Rightarrow \quad \hat{y}_i = \bar{y} + w(x_i - \bar{x}) \qquad (\because\ b = \bar{y} - w\bar{x}\ )$$

$$\Rightarrow \quad (\hat{y}_i - \bar{y})^2 = w^2(x_i - \bar{x})^2$$

$$\Rightarrow \quad \sum_{i=1}^{m}(\hat{y}_i - \bar{y})^2 = w^2\sum_{i=1}^{m}(x_i - \bar{x})^2$$

Hence we have

$$\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{m}(y_i - \bar{y})^2 - \sum_{i=1}^{m}(\hat{y}_i - \bar{y})^2 \ .$$

Linear Regression

We call the terms as follows

- Residual Sum of Squares, $RSS = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$
- Total Sum of Squares, $TSS = \sum_{i=1}^{m} (y_i - \overline{y})^2$
- Explained Sum of Squares, $ESS = \sum_{i=1}^{m} (\hat{y}_i - \overline{y})^2$

In other words, we have

$$RSS = TSS - ESS \ .$$

Coefficient of determination

The *coefficient of determination* (決定係數) is defined as

$$R^2 = \frac{ESS}{TSS}$$

and thus

$$RSS = (1 - R^2)TSS\ .$$

Coefficient of determination

$$\because \quad R^2 = \frac{ESS}{TSS}$$

$$= \frac{\sum_{i=1}^{m} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{m} (y_i - \bar{y})^2}$$

$$= \frac{w^2 \sum_{i=1}^{m} (x_i - \bar{x})^2}{\sum_{i=1}^{m} (y_i - \bar{y})^2}$$

$$= \frac{\left[\sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\left[\sum_{i=1}^{m} (x_i - \bar{x})^2\right]\left[\sum_{i=1}^{m} (y_i - \bar{y})^2\right]} \qquad \left(\because w = \frac{\sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m} (x_i - \bar{x})^2} \right)$$

$$= \frac{\left[Cov(X, Y)\right]^2}{Cov(X, X)Cov(Y, Y)}$$

we have

$$R^2 = \rho_{X,Y}{}^2$$

where $\rho_{X,Y} = \dfrac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}$ is the *Pearson correlation coefficient*.

Summary

$$\hat{y}_i = wx_i + b$$

$$\overline{y} = w\overline{x} + b$$

$$w = \frac{Cov(X, Y)}{Cov(X, X)}$$

$$RSS = TSS - ESS = (1 - R^2)TSS$$

where $R^2 = {\rho_{X,Y}}^2$: coefficient of determination

$\rho_{X,Y} = \dfrac{Cov(X,Y)}{\sigma_X \sigma_Y}$: Pearson correlation coefficient

$RSS = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$: residual sum of squares

$TSS = \sum_{i=1}^{m} (y_i - \overline{y})^2$: total sum of squares

$ESS = \sum_{i=1}^{m} (\hat{y}_i - \overline{y})^2$: explained sum of squares

Note the *mean square error*, $\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m} (\hat{y}_i - y_i)^2$.