How do we obtain the values of $\boldsymbol{\beta} = [\beta_0, \beta_1 \, \ldots \, \beta_{p-1}]^T$

$$\boldsymbol{X} = \begin{bmatrix} x_{00} & x_{01} & \ldots & x_{0p-1} \\ \vdots & & & \\ x_{n-10} & \ldots & \ldots & x_{n-1p-1} \end{bmatrix}$$

$$\boxed{y \to y(\boldsymbol{x}) = f(\boldsymbol{x}) + \varepsilon}$$

$$\varepsilon \sim N(0, \sigma^2)$$

mean value $\mu = 0$

iid = independent and identically distributed

$$f(x) \simeq \tilde{y}(x) \to \tilde{y}_i = \tilde{y}(x_i)$$

$$y = X\beta + \varepsilon \qquad \tilde{y} = X\beta$$

How do we find β?
optimizing/minimizing a function called the cost/loss function

In linear regression:

$$\text{Cost function} = C(\beta; X)$$

$$= \frac{1}{} \sum^{n-1} (y_i - \tilde{y}_i)^2 = MSE$$

$$n \qquad i=0$$

$$= \frac{1}{n}\left[(y-\tilde{y})^T(y-\tilde{y})\right]$$

$$= \frac{1}{n}\left[(y-X\beta)^T(y-X\beta)\right]$$

$$\frac{1}{n}\sum_{i=0}^{n-1}\left(y_i - X_{i,*}\beta\right)^2$$

$$\min_{\beta\in\mathbb{R}^p}\left[(y-X\beta)^T(y-X\beta)\right]$$

$$\frac{\partial C(\beta)}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}\left[\sum_{i=0}^{n-1}\left(y_i - \beta_0 x_{i0} - \beta_1 x_{i1} - \beta_2 x_{i2} \dots \beta_j x_{ij} - \dots - \beta_{p-1} x_{i\,p-1}\right)^2\right.$$

$$X_{i*}\beta = x_{i0}\beta_0 + x_{i1}\beta_1 + \dots x_{i\,p-1}\beta_{p-1}$$

$$\boxed{\begin{bmatrix} x_{00} & x_{01} & \dots & x_{0\,p-1} \\ x_{10} & x_{11} & \dots & x_{1\,p-1} \\ \vdots & & & \\ x_{1000\,0} & \dots & \dots & x_{1000\,p-1} \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{bmatrix}}$$

$$\frac{\partial C}{\partial \beta_j} = 0 = -2\sum_{i=0}^{n-1} \underline{x_{ij}}\left(y_i - \beta_0 x_{i0} - \beta_1 x_{i1}\right.$$

$$\vdots$$

$$- \beta_{p-1} x_{i,p-1})$$

$\Rightarrow$ in Matrix vector form ;

$$\frac{\partial C}{\partial \beta} = 0 = X^T (y - X\beta)$$

$$\Rightarrow \quad X^T X \beta = \underline{X^T y}$$

$$X \in \mathbb{R}^{n \times p} \qquad X^T \in \mathbb{R}^{p \times n}$$

$$y \in \mathbb{R}^n$$

if we can invert $X^T X$

$$\Rightarrow \quad \beta = \underbrace{(X^T X)^{-1} X^T y}_{\text{known.}}$$