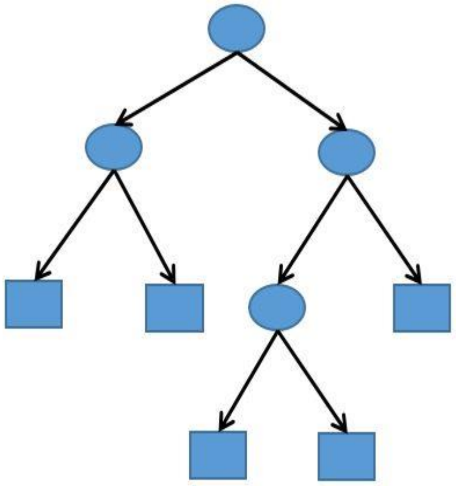
法律声明

关注 小象学院

Machine Learning

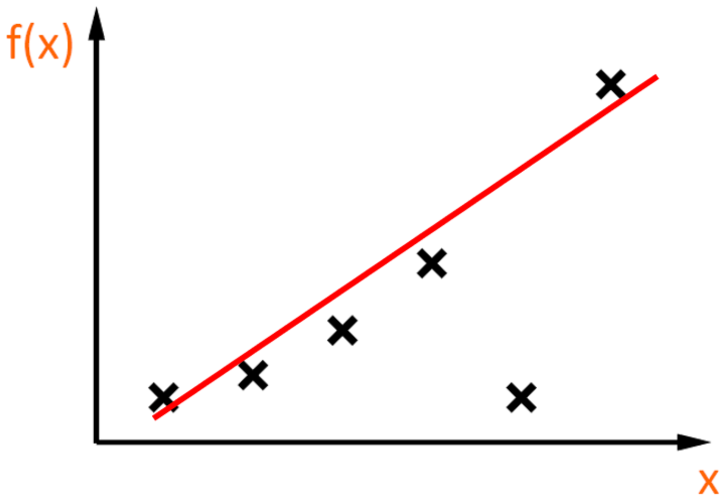
Part 4: Classical Machine Learning Models

Zengchang Qin (Ph.D.)

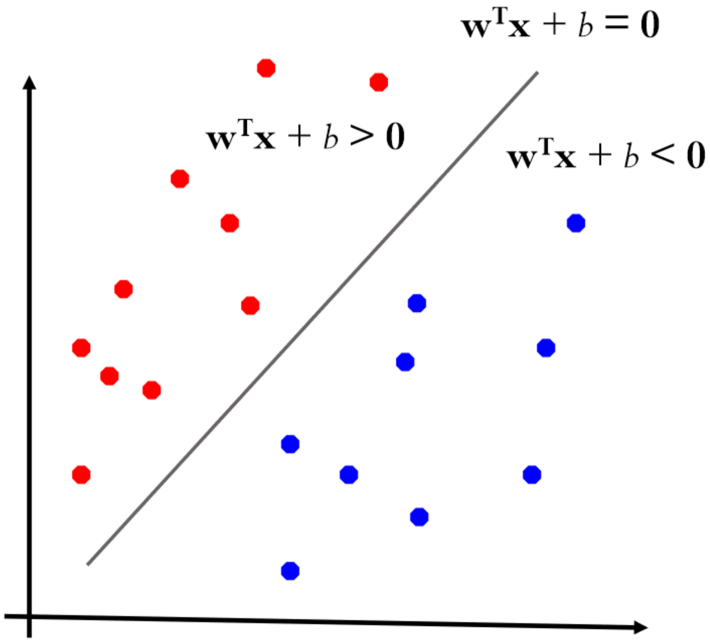


Linear Model

Linear Model

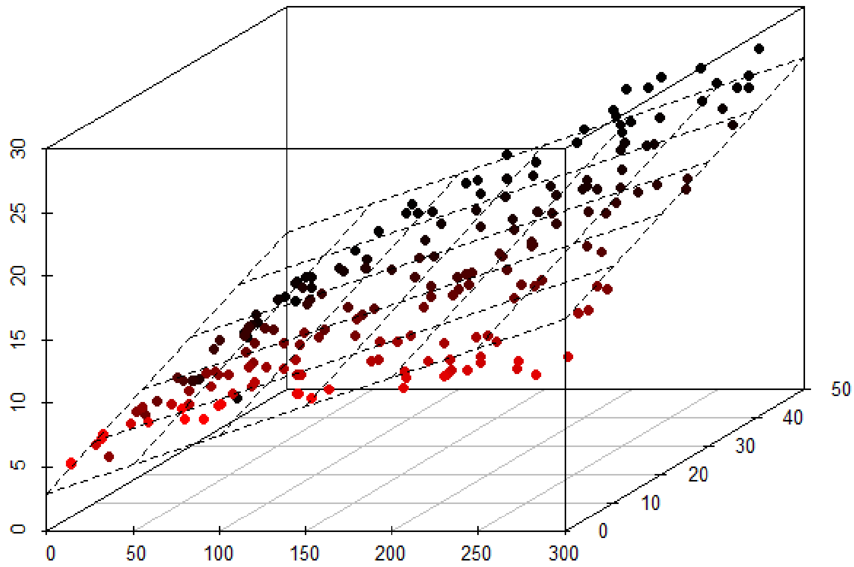


Linear fit



Linear decision boundary

Linear Regression



$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x$$

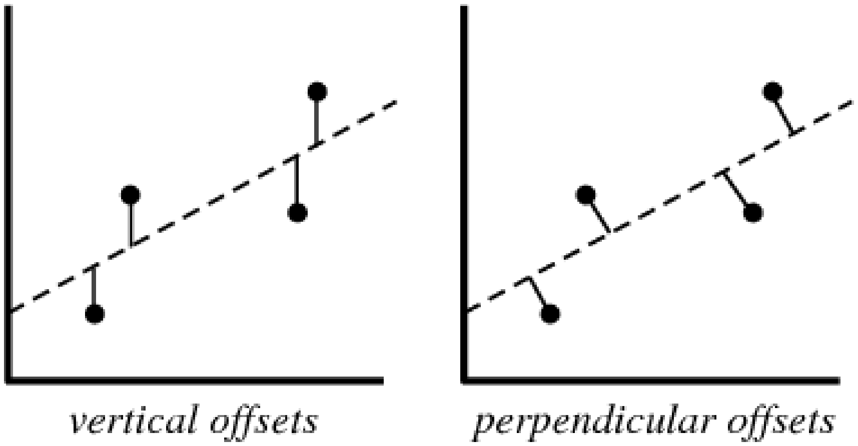
$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

In statistics, the most common occurrence is in connection with regression models and the term is often taken as synonymous with linear regression model.

Least Square

A mathematical procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets.

$$S = \sum_{i=1}^{n} {r_i}^2 \qquad r_i = y_i - f(x_i, \boldsymbol{\beta}).$$



Solving the least squares problem:

$$\frac{\partial S}{\partial \beta_j} = 2\sum_i r_i \frac{\partial r_i}{\partial \beta_j} = 0,\ j = 1, \ldots, m,$$

$$= -2\sum_i r_i \frac{\partial f(x_i, \boldsymbol{\beta})}{\partial \beta_j} = 0,\ j = 1, \ldots, m.$$

since $\quad r_i = y_i - f(x_i, \boldsymbol{\beta})$

Residuals are the vertical distances between the data points and the corresponding predicted values.

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y}.$$

Matrix

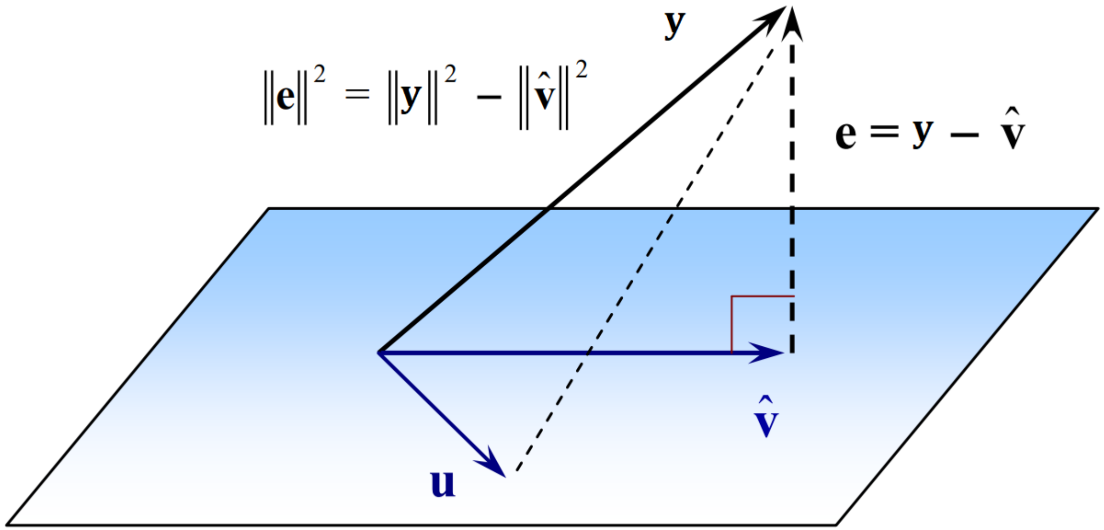
$$\sum_{j=1}^{n} X_{ij}\beta_j = y_i,\ (i = 1, 2, \ldots, m), \quad \mathbf{X}\boldsymbol{\beta} = \mathbf{y},$$

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} S(\boldsymbol{\beta}), \quad S(\boldsymbol{\beta}) = \sum_{i=1}^{m} \left| y_i - \sum_{j=1}^{n} X_{ij}\beta_j \right|^2 = \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^2$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^{\mathrm{T}}\mathbf{y}.$$

Geometrical Interpretation



$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}.$$

Probabilistic Interpretation

- Let us assume that the target variables and the inputs are related via the equation:

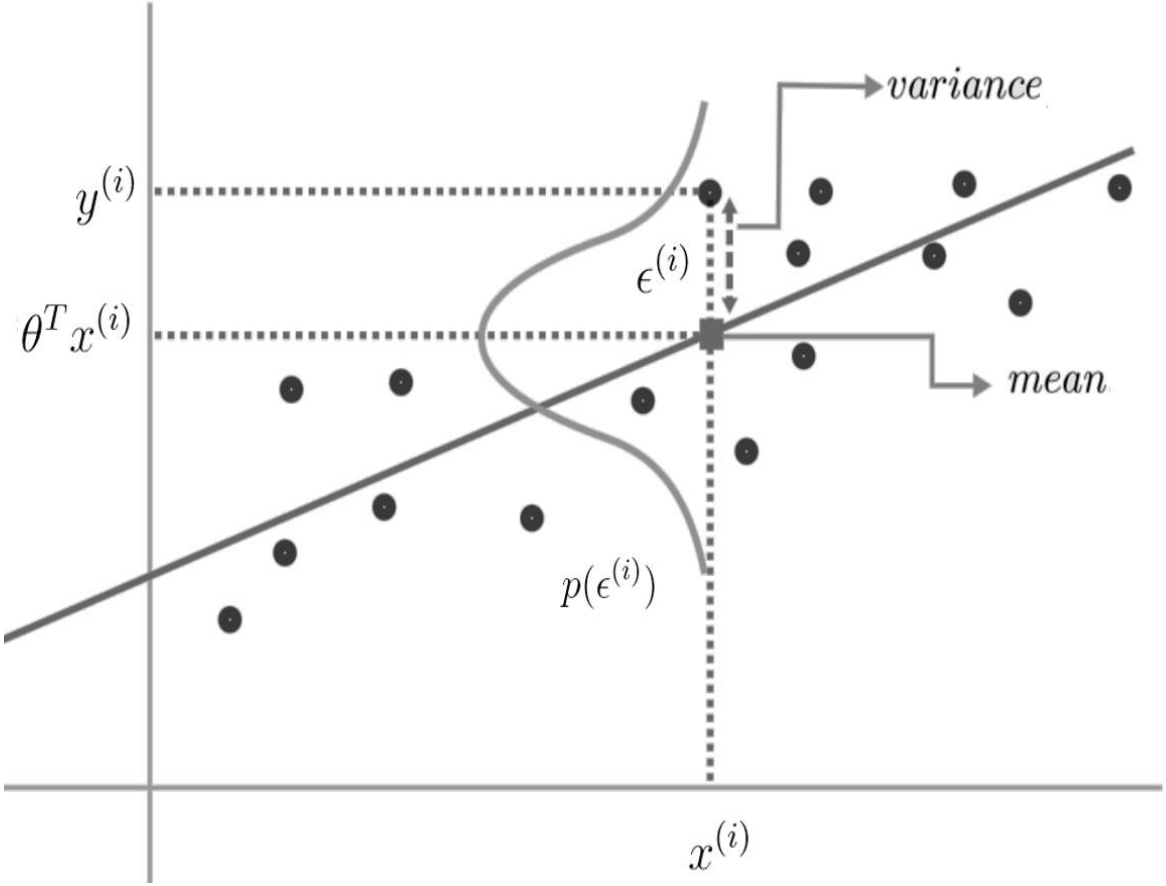
$$\boxed{y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}}$$

where $\varepsilon^{(i)}$ is an error term.

- We can write this assumption as $\varepsilon^{(i)}$ follows a Normal distribution:

$$\boxed{p(\epsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\epsilon^{(i)})^2}{2\sigma^2}\right)}$$

Likelihood and Cost Function



Likelihood

$X_1, X_2, X_3, \ldots X_n$ have joint density denoted

$$f_\theta(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n|\theta)$$

Given observed values $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, the likelihood of θ is the function

If $X_1, X_2, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$ random variables their density is written:

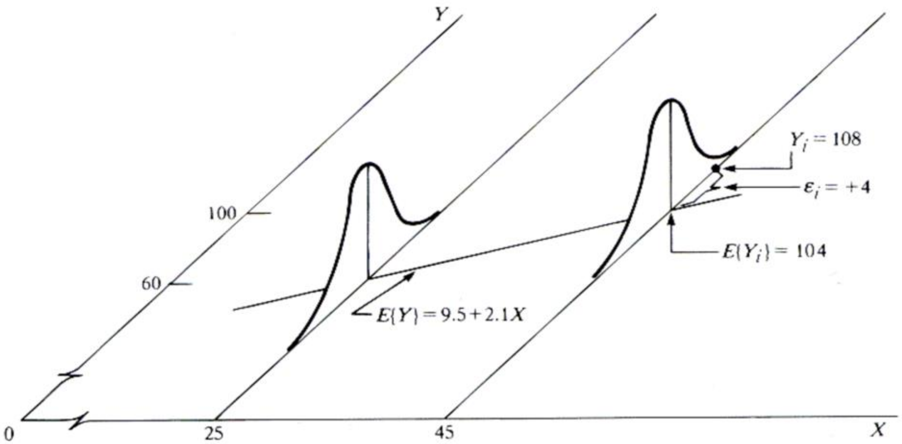
$$f(x_1, \ldots, x_n|\mu, \sigma) = \prod_i^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}[\frac{x_i - \mu}{\sigma}]^2\right)$$

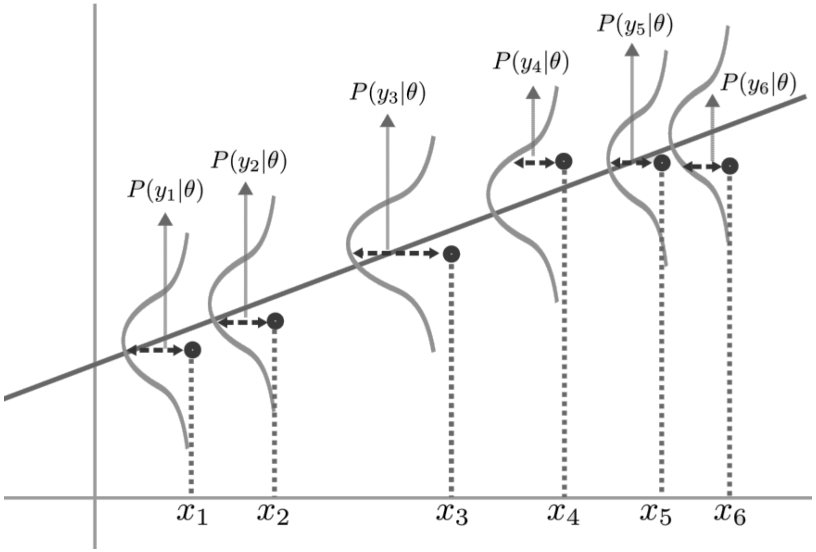
$$f(x_1, x_2, \ldots, x_n \mid \theta) = f(x_1 \mid \theta) \times f(x_2 \mid \theta) \times \cdots \times f(x_n \mid \theta).$$

$$\mathcal{L}(\theta\,; x_1, \ldots, x_n) = f(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^n f(x_i \mid \theta).$$

$$\boxed{\ln \mathcal{L}(\theta\,; x_1, \ldots, x_n) = \sum_{i=1}^n \ln f(x_i \mid \theta),}$$

ML Estimation



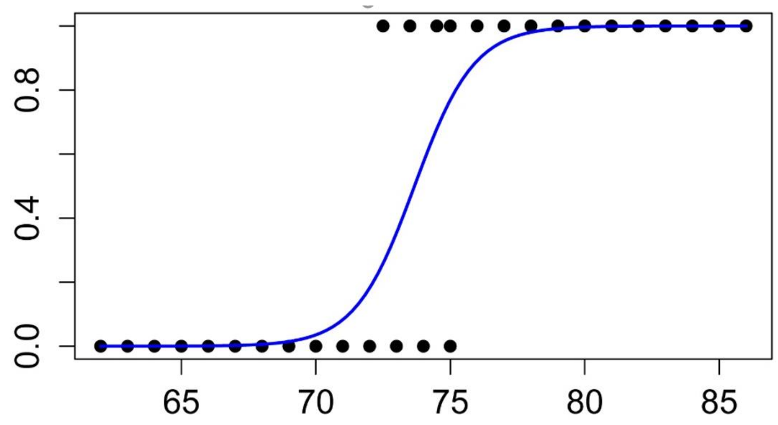


It implies that: $p(y^{(i)}|x^{(i)};\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$

- When we wish to explicitly view this as a function of θ, we will instead call it the likelihood function:

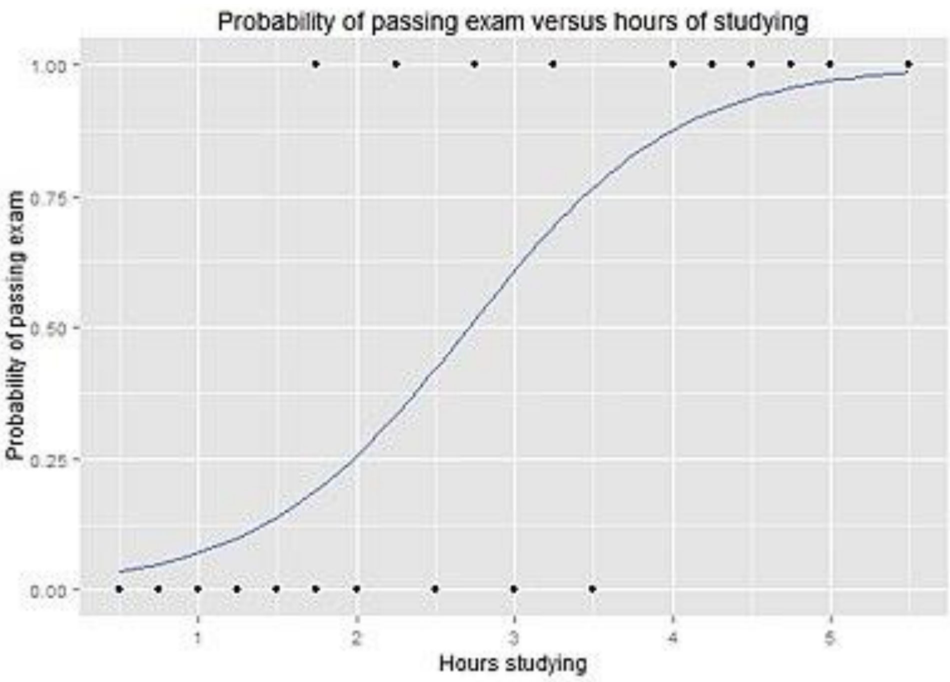
$$\boxed{L(\theta) = L(\theta; X, \vec{y}) = p(\vec{y}|X;\theta) = \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)};\theta) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)}$$

Logistic Regression

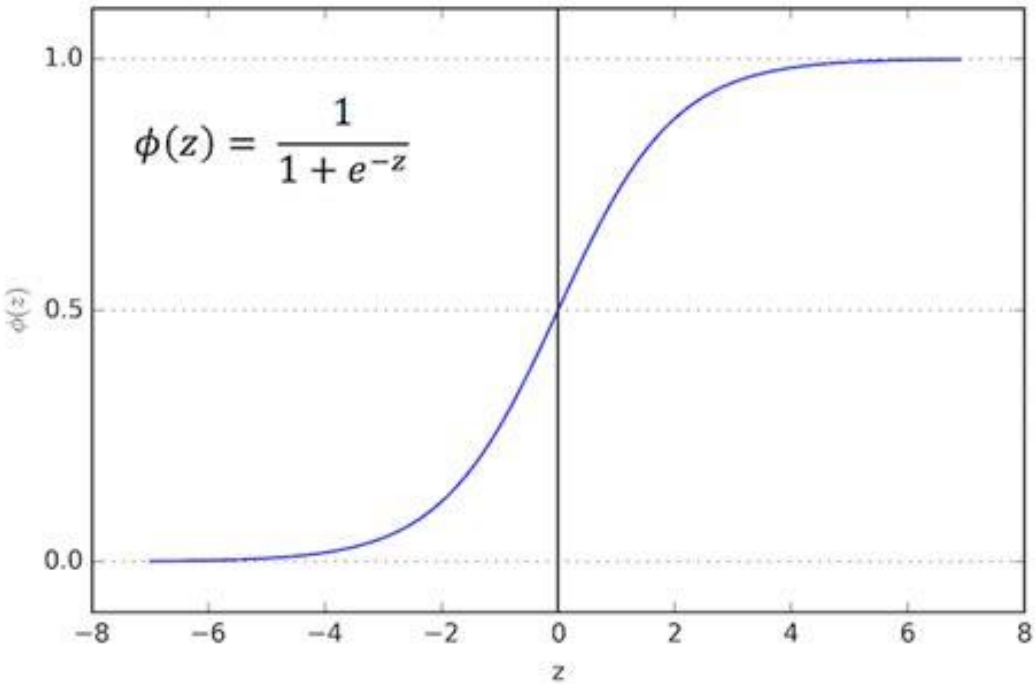


A Simple Classification Problem

Hours	0.5	0.75	1	1.25	1.50	1.75	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Pass	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1



Sigmoid Function



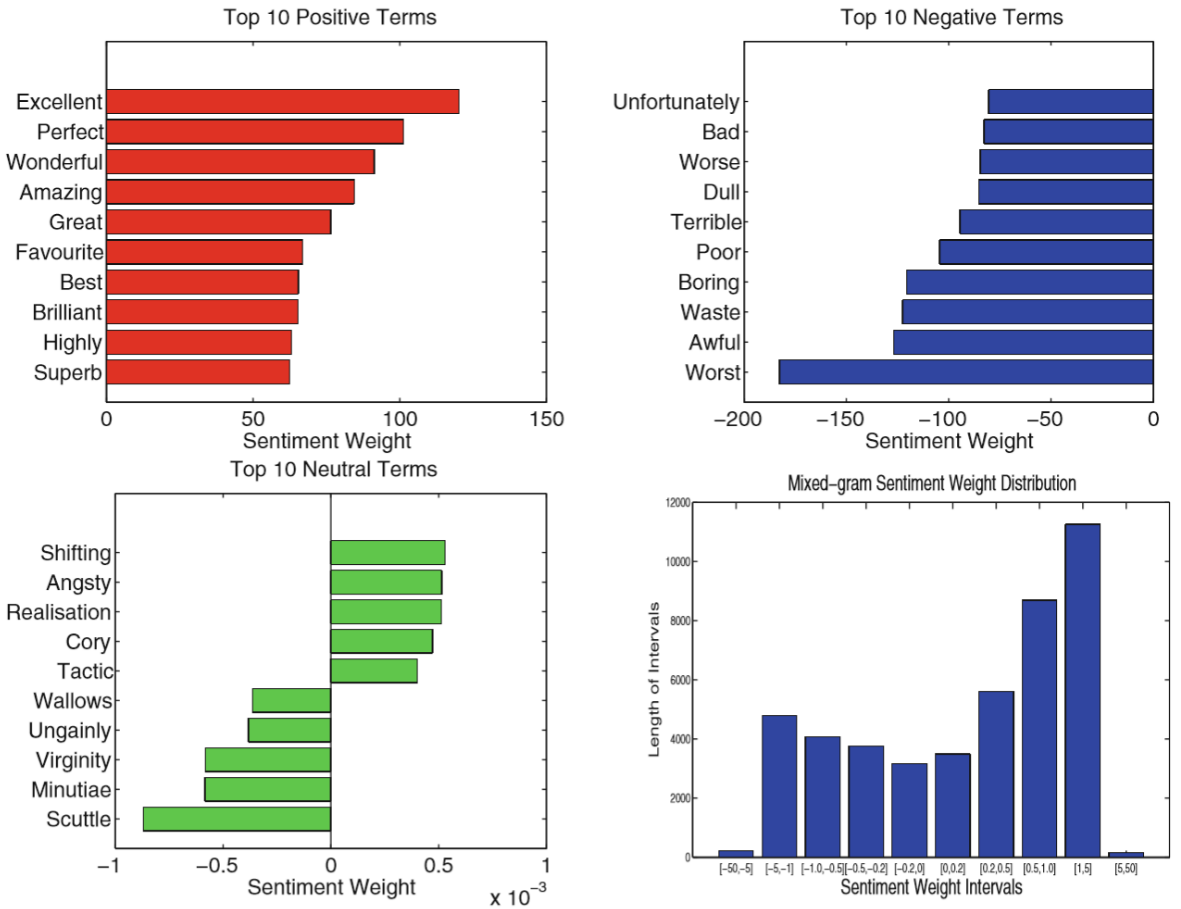
Equations

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}^\top \mathbf{x})}$$

- How to learn the parameters?
- Can we do MLE?

Application

$$h = f\left(\sum_{i=1}^{N} w_i x_i\right) = f(\mathbf{w}^T \mathbf{x}) \qquad h_j = \frac{1}{1 + \exp\left(-\mathbf{w}^T \mathbf{x}_j\right)} \quad ; \quad 1 \le j \le M$$



联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：**小象学院**