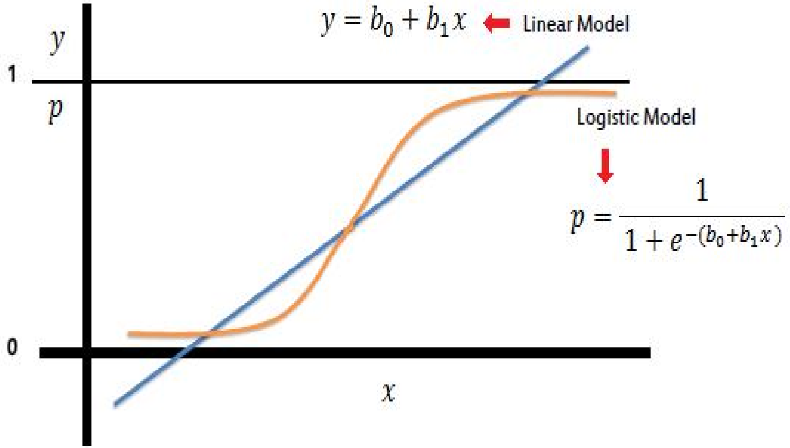
Logistic Regression Muti-class in text classification

Bui Thanh Phong - 51600063

1. logistic regression?

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary

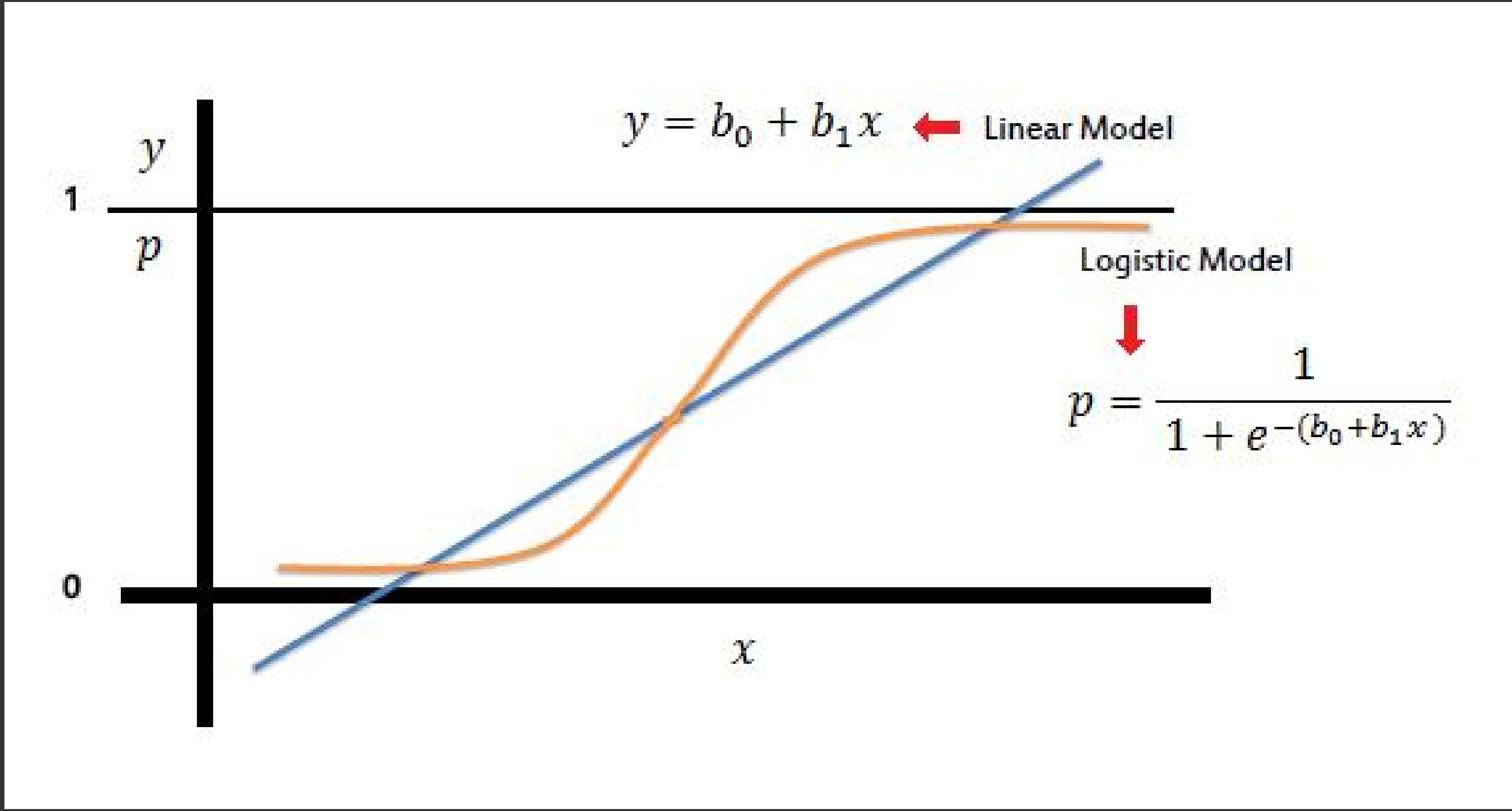
Like all regression analyses, the logistic regression is a predictive analysis.
Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.



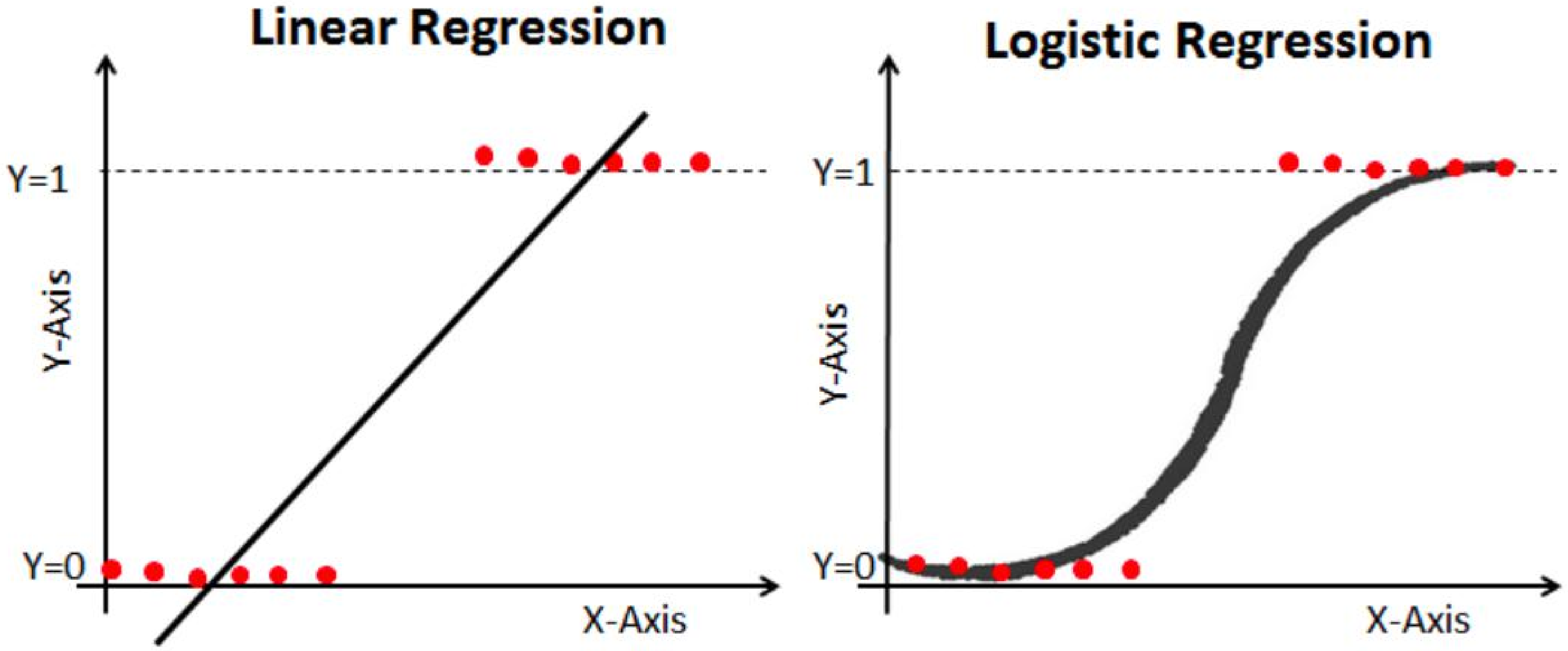
Intro

- → Logistic Regression Model
- → Sigmoid function
- → Loss function
- → Stochastic Gradient Descent
- → Logistic Sigmoid Regression
- → Demo

– Logistic Regression Model

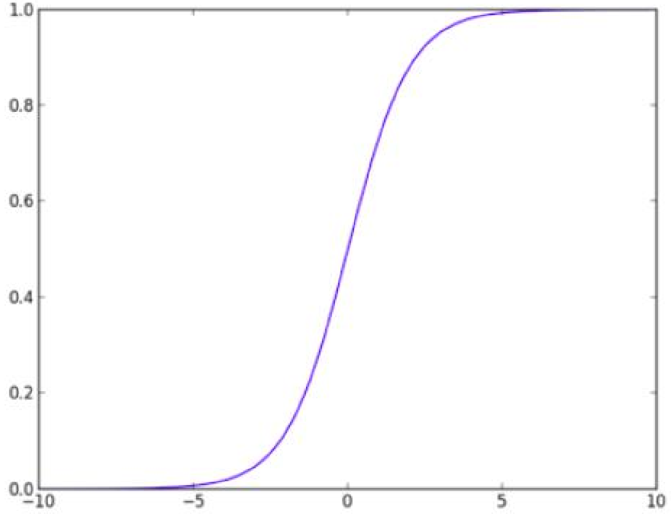


– Logistic Regression Model



- Sigmoid function

$$f(x) = \frac{1}{1 + e^{-(x)}}$$



Sigmoid function $f(s) = \frac{1}{1+e^{-s}}$

$$f'(s) = \frac{e^{-s}}{(1+e^{-s})^2} = \frac{1}{1+e^{-s}} . \frac{e^{-s}}{1+e^{-s}} = \frac{1}{1+e^{-s}} . \frac{e^{-s}+1-1}{1+e^{-s}}$$

$$= \frac{1}{1+e^{-s}} . \left(\frac{e^{-s}+1}{1+e^{-s}} - \frac{1}{1+e^{-s}} \right)$$

$$= f(s) . (1 - f(s))$$

$$\lim_{s \to -\infty} f(s) = 0; \quad \lim_{s \to +\infty} f(s) = 1$$

Loss function

$$P(y_i = 1|\mathbf{x}_i; \mathbf{w}) = \qquad f(\mathbf{w}^T\mathbf{x}_i) \quad (1)$$

$$P(y_i = 0|\mathbf{x}_i; \mathbf{w}) = 1 - f(\mathbf{w}^T\mathbf{x}_i) \quad (2)$$

$$z_i = f(\mathbf{w}^T\mathbf{x}_i)$$

$$P(y_i|\mathbf{x}_i; \mathbf{w}) = z_i^{y_i}(1 - z_i)^{1-y_i}$$

Loss function

$$P(\mathbf{y}|\mathbf{X};\mathbf{w}) = \prod_{i=1}^{N} P(y_i|\mathbf{x}_i;\mathbf{w})$$

$$= \prod_{i=1}^{N} z_i^{y_i}(1 - z_i)^{1-y_i}$$

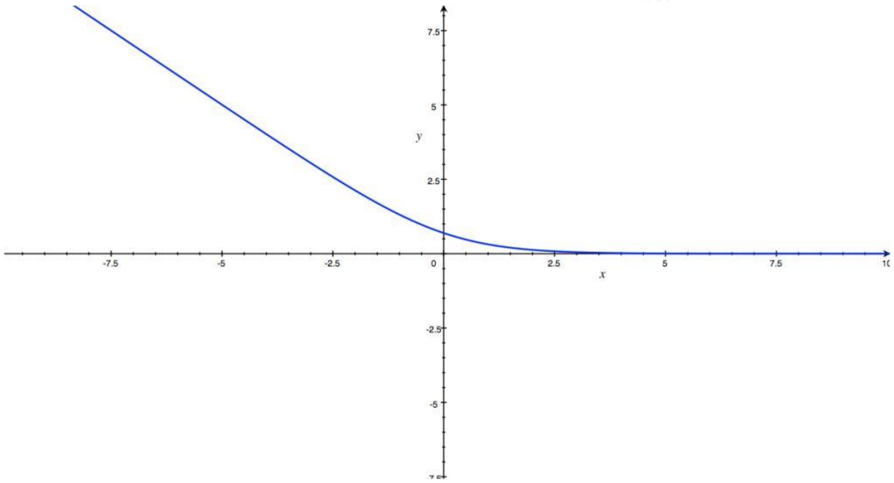
Loss function

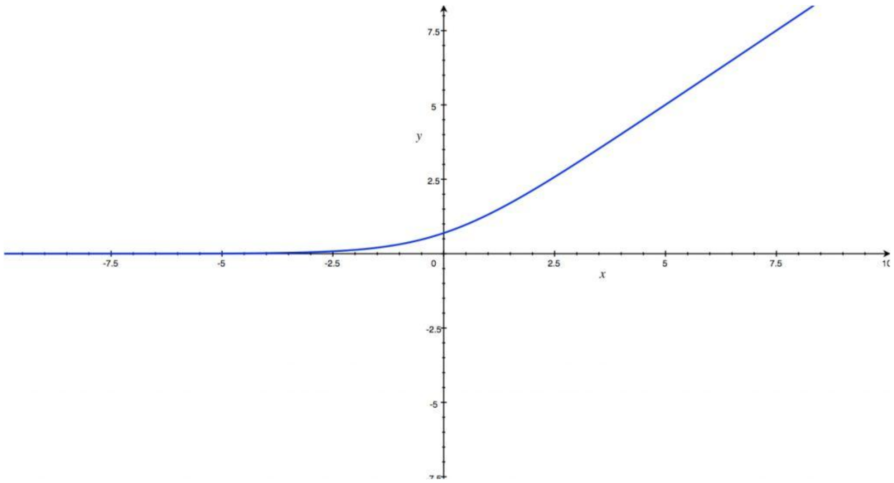
$$J(\mathbf{w}) = -\log P(\mathbf{y}|\mathbf{X};\mathbf{w})$$

$$= -\sum_{i=1}^{N}(y_i \log z_i + (1 - y_i)\log(1 - z_i))$$

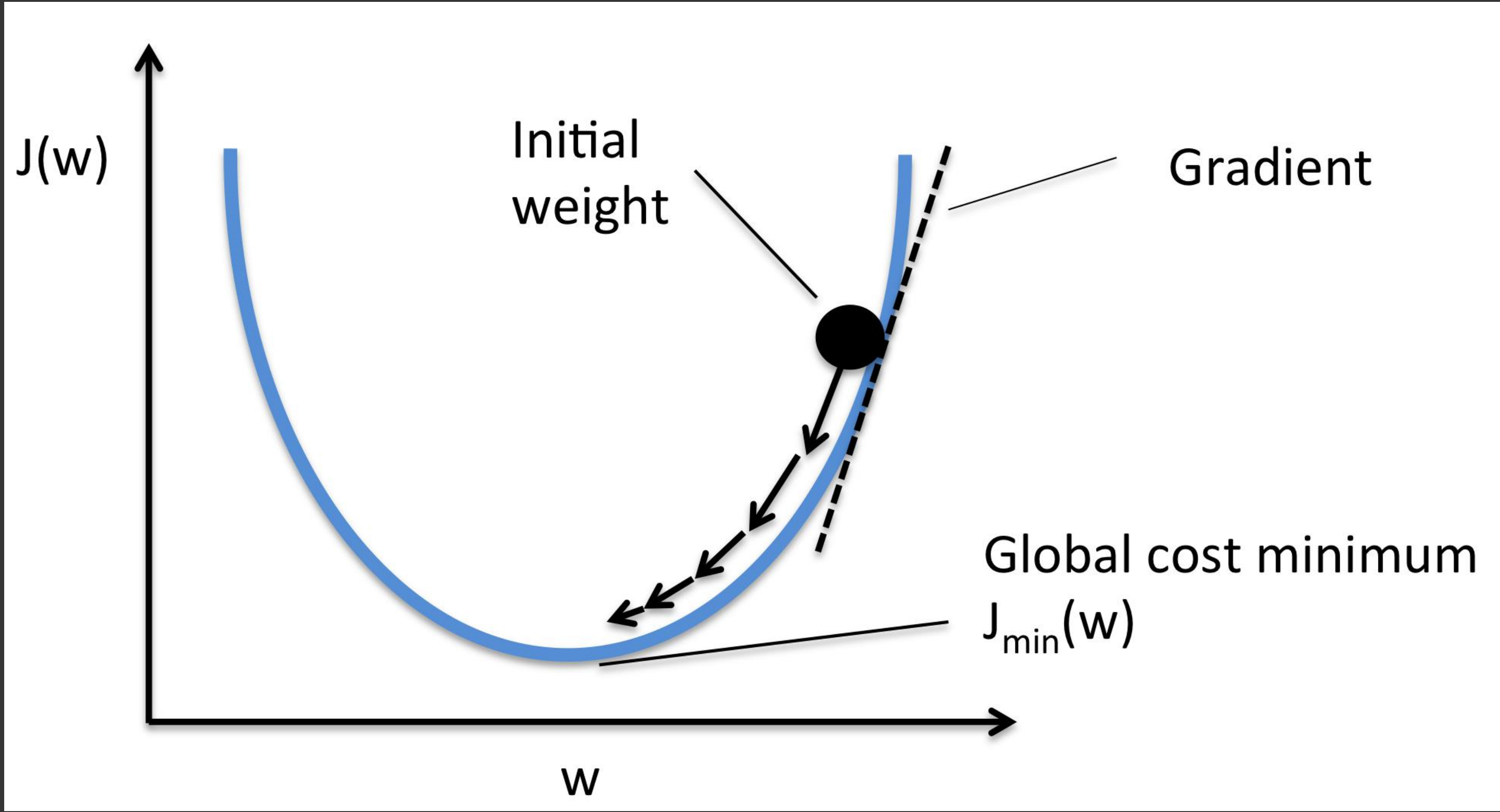
Loss function

$$J(\theta) = \begin{cases} - \log(h(x) & (\text{if } y \ = 1) \\ - \log(1 - (h(x)) & (\text{if } y \ = 0) \end{cases}$$





Stochastic Gradient Descent



Stochastic Gradient Descent

$$J(\mathbf{w}; \mathbf{x}_i, y_i) = -(y_i \log z_i + (1 - y_i) \log(1 - z_i))$$

$$\begin{aligned}\frac{\partial J(\mathbf{w}; \mathbf{x}_i, y_i)}{\partial \mathbf{w}} &= -(\frac{y_i}{z_i} - \frac{1 - y_i}{1 - z_i}) \frac{\partial z_i}{\partial \mathbf{w}} \\ &= \frac{z_i - y_i}{z_i(1 - z_i)} \frac{\partial z_i}{\partial \mathbf{w}}\end{aligned}$$

Stochastic Gradient Descent

Chain rule: $\dfrac{\partial z_i}{\partial \mathbf{w}} = \dfrac{\partial z_i}{\partial s}\dfrac{\partial s}{\partial \mathbf{w}} = \dfrac{\partial z_i}{\partial s}\mathbf{x}$

$$\frac{\partial J(\mathbf{w};\mathbf{x}_i,y_i)}{\partial \mathbf{w}} = (z_i - y_i)\mathbf{x}_i$$

$$\mathbf{w} = \mathbf{w} + \eta(y_i - z_i)\mathbf{x}_i$$

2. Demo

http://viet.jnlp.org

Maximum Entropy Model

feature

$$\frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} f_i(d, c(d)) = \sum_{d} \mathrm{P}(d) \sum_{c} \mathrm{P}(c|d) f_i(d, c). \quad (1)$$

$$\frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} f_i(d, c(d)) = \frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} \sum_{c} \mathrm{P}(c|d) f_i(d, c). \quad (2))$$

$$f_{w,c'}(d, c) = \begin{cases} 0 & \text{if } c \neq c' \\ \frac{N(d,w)}{N(d)} & \text{Otherwise,} \end{cases}$$