Classification (Binary)

Friday, 25 October 2024 11:06

$y \in \mathbb{R}$

$y = \{0, 1\}$ → 1: Positive outcome, 0: Negative Outcome

$y = \{0, 1, 2\}$ → 0: Not Spam, 1: Unsure, 2: Spam

C = Symbol to denote the cardinality of classes

$\hat{y} = f_\theta(x) = h_\theta(x)$

$\hat{y}$: belongs to probability simplex Δ_C

$y_i \geq 0$ → single element

$\sum_i y_i = 1$

$x_i = $ (0 → 0.7, 1 → 0.3)

$f_\theta(x) = \hat{y} \in \Delta_C$ we can interpret $\hat{y}$ as a categorical distribution

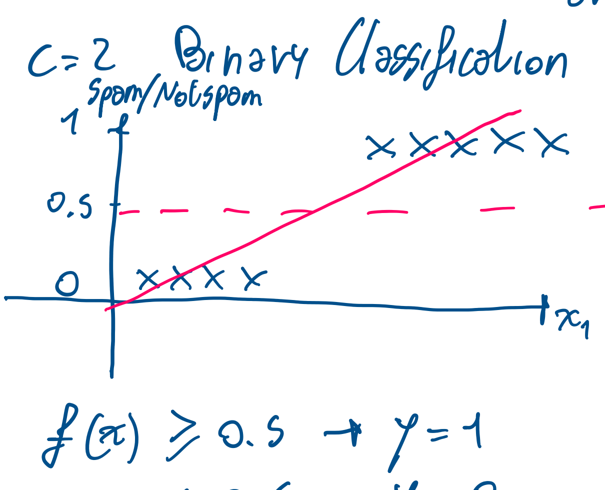
$c := \underset{i}{\text{argmax}}\ \hat{y}$

$[0, 1]$

$[0.7, 0.3]$

$\text{argmax}\ [\checkmark, \times] \to$ class 0

C=2 Binary Classification



lines word credits is used

$f(x) \geq 0.5 \to y = 1$

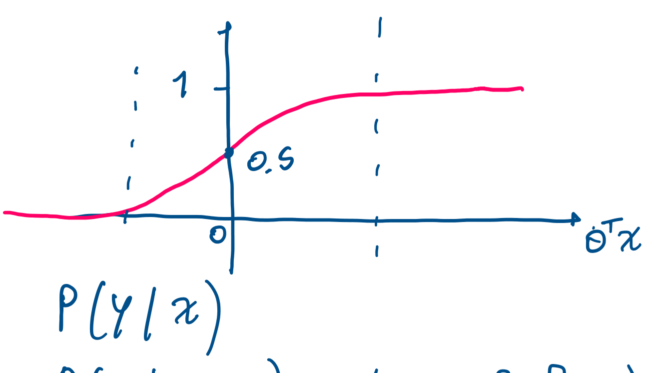
$< 0.5 \to y = 0$

$f_\theta(x) = \pm\infty$

let apply a function $g(f_\theta(x)) \to \Delta_C$

$g(\cdot) = \text{Sigmoid function} = \sigma(\theta^T x_i) =$

$$= \frac{1}{1 + e^{-\theta^T x}}$$

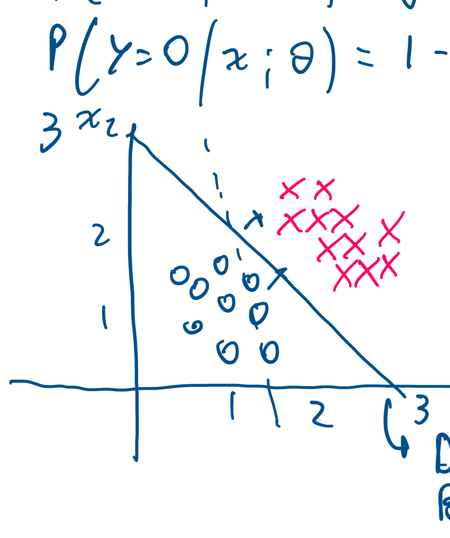


$P(y \mid x)$

$P(y \mid x; \theta)$ alcore of $f_\theta(x)$

$P(y = 1 \mid x; \theta) = f_\theta(x)$

$P(y = 0 \mid x; \theta) = 1 - f_\theta(x)$

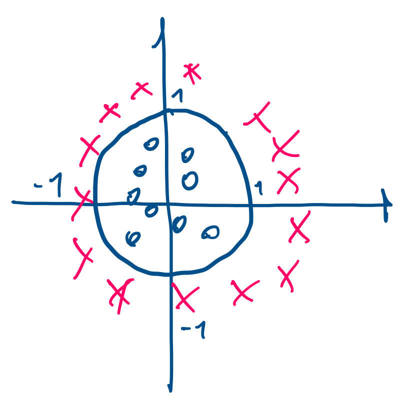


$f_\theta(x) =$

$= \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

$= \sigma(3 + (-1 x_1) + (-1 x_2))$

Predict if $-3 + x_1 + x_2 \geq 0.5$



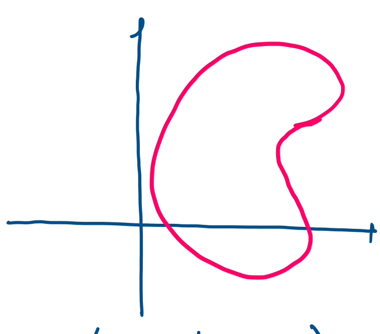
Non Linear Decision Boundry.

$f_\theta(x) = [\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2]$

$\theta_0 = -1,\ \theta_1 = 0,\ \theta_2 = 0,\ \theta_3 = 1,\ \theta_4 = 1$

$-1 + x_1^2 + x_2^2 \geq 0 \to y = 1$

$x_1^2 + x_1^2 = 1$ $\quad < 0 \to y = 0$



Bean decision Boundary?

$\partial_*(x_1 x_2^2)$

$P(y_i = 1 \mid x_i; \theta) = f_\theta(x)$

$P(y_i = 0 \mid x_i; \theta) = 1 - f_\theta(x)$

$$P(y_i \mid x_i; \theta) = \begin{cases} f_\theta(x_i) & \text{if } y_i = 1 \\ 1 - f_\theta(x_i) & \text{if } y_i = 0 \end{cases} = f_\theta(x_i)^y \left(1 - f_\theta(x_i)\right)^{1-y}$$

$$P(\underbrace{y}_{\text{vector}} \mid \underbrace{X}_{\text{Matrix}}; \theta) = \prod^n f_\theta(x_i)^{y_i} \left(1 - f_\theta(x_i)\right)^{1-y_i}$$

↳ Likelihood

Log Likelihood $\log L(\theta) = \frac{1}{n} \sum^n y_i \log f_\theta(x) + (1 - y_i) \log(1 - f_\theta(x))$

In practice we minimize the Negative Log Likelihood

Max Likelihood = Min NLL

From a numerical point of view ensures that as the probability for the right class increases the loss decreases

$y_i = 1 \quad \log(\underbrace{\sigma(\theta^T x_i)}_{\text{close to } 1})$

$$NLL(\theta) = -\frac{1}{n} \sum^n y_i \log f_\theta(x) + (1 - y_i) \log(1 - f_\theta(x))$$

$y_i \log(\sigma(\theta^T x))$

↓ Chain Rule $\quad \frac{\partial \log(\sigma(z))}{\sigma(z)} = \frac{1}{\sigma(z)} \cdot \sigma'(z)$

$\frac{\partial \sigma(z)}{\partial z} = \sigma(z)(1 - \sigma(z))$

$\frac{\partial \theta^T x_i}{\partial} = x_i$

$$\frac{\partial\, y_i \log(\sigma(\theta^T x))}{\partial \theta} = y_i \cdot \frac{1}{\sigma(\theta^T x)} (\sigma(\theta^T x_i))(1 - \sigma(\theta^T x)) x_i$$

$$= y_i (1 - \sigma(\theta^T x_i)) x_i$$

Chain Rule $\quad \frac{\partial \log(1 - \sigma(z))}{\sigma(z)} = \frac{-1}{(1 - \sigma(z))} \cdot \sigma'(z)$

$$= (1 - y_i) \frac{-1}{(1 - \sigma(\theta^T x_i))} \sigma(\theta^T x_i)(1 - \sigma(\theta^T x_i)) x_i$$

$$= -(1 - y_i) \sigma(\theta^T x) x_i$$

$$\frac{\partial NLL(\theta)}{\partial \theta} = -\frac{1}{n} \sum^n (y_i - \sigma(\theta^T x_i)) x_i$$

↓ difference between Real Lable and the predicted one

$$\theta \leftarrow \theta + \alpha \frac{1}{n} \sum^n (y_i - \sigma(\theta^T x_i)) x_i$$

↓ if error is large → the update will be large

" " is small → " " " " small

$(0 - 0.3) \cdot x_i$

$(1 - 0.7) \cdot x_i$

$(0 - 0.7) x_i$

Replaced by Newton's Method

$x \quad x'$ velocity $\quad x''$ acceleration

x': how quickly (change) in which direction

x'': rate of change of first derivative

Use this information to adjust the step size

$\theta \leftarrow \theta - H^{-1} \nabla NLL(\theta)$

$H = \sigma(\theta^T x_i)(1 - \sigma(\theta^T x_i)) x_i x_i^T$

↳ H^{-1} is the invers → Costly and not alway guaranteed

Algoritm BFGS: it is a Quasi-Newton Method

↳ Compute approximation of the Hessian.