Supervised Learning ⇒

Dataset ⇒ Features ⇒ Input Variable

Labels / targets ⇒ Prediction

Regression

Quantitative

$\in \mathbb{R}$

Price sales estimate.

Classification

Qualititative / Discrete

$\in S = \{s_1, s_2 \ldots s_n\}$ 3/5/10/12

2-classes ⇒ True / False ⇒ Binary classification

3- " ⇒

n- "

>2 ⇒ multinomial classification

Regression $\Rightarrow$ Label $\in \mathbb{R}$

Linear Regression $\Rightarrow$

Single Lin. Reg.

$\downarrow$

1-Feature

Multiple Lin. Reg.

$\downarrow$

> 1-Feature

10

Polynomial Lin. Reg.

$\boxed{y = mx + C}$ $\Rightarrow$ Straight Line

$h_\theta(x) \Rightarrow y = \theta_0 + \theta_1 x_1 \Rightarrow \beta_0 + \beta_1 x_1$

y: Dependent Variable

θ_0: Intercept

θ_1: Slope

x_1: Independent Variable

$h_\theta(x) = \theta_0 + \theta(x) \Rightarrow \hat{y} \Rightarrow$ Prediction

$\theta_0 \cdot 1 = \theta_0 x_0 = 1$

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n \Rightarrow$ where $\Rightarrow n$ is number of features

$$= \theta X^T \Rightarrow \theta = [\theta_0 \; \theta_1 \; \theta_2 \dots \theta_n]_{1 \times n} \Rightarrow \text{Parameters / slopes}$$

$$= \theta^T X \qquad X = [1, x_1 \; x_2 \dots x_n]_{1 \times n} \Rightarrow \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \Rightarrow X^T$$

$$\sum_{i=0}^{n} \theta_i x_i \Rightarrow x_0 = 1$$

Objectives $\Rightarrow$

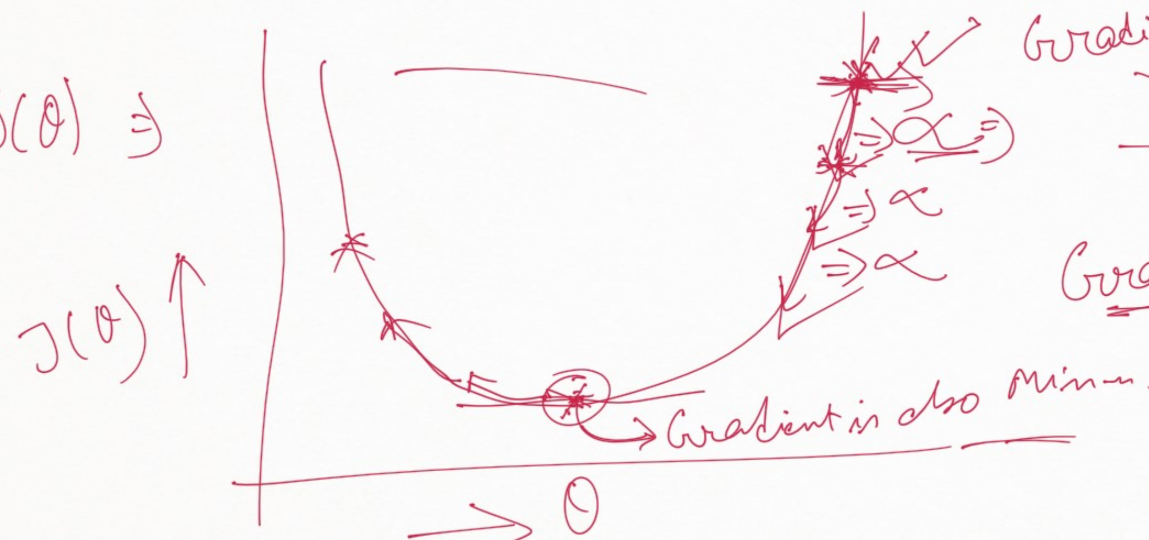
i) Find out optimal $\theta \Rightarrow [\theta_0 \; \theta_1 \; - - \; \theta_n]$

ii) Measure Accuracy of f

i) Cost Function $\Rightarrow J(\theta) \Rightarrow$ Mean Residual Sum of Squares

$$\Rightarrow \frac{\sum_{i=1}^{n} \left(h_\theta(x_i) - y_i\right)^2}{n} \Rightarrow \frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n}$$

$J(\theta) \Rightarrow$



Gradient of a function ⇒

$$\frac{\partial (J(\theta))}{\partial \theta}$$

Gradient Descent

$$\theta_j := \theta_j - \alpha \cdot \frac{\partial}{\partial \theta}(J(\theta))$$

Repeat until Convergence

$$\frac{\partial (J(\theta))}{\partial \theta} = \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \left(h_\theta(x_i) - y_i \right)^2 \Rightarrow$$

$$\theta = (X^T X)^{-1} \cancel{Y^T} X$$

$$J(\theta) = \sum_{i=1}^{n} \left(h_\theta(x_i) - y_i\right)^2$$

$$= \left(h_\theta(x_i) - y_i\right)^T \left(h_\theta(x_i) - y_i\right)$$

$$= (X\theta - Y)^T (X\theta - Y)$$

$$= (X^T\theta^T - Y^T)(X\theta - Y)$$

$$= X^T\theta^T X\theta - YX^T\theta^T - Y^T X\theta + YY^T$$

$$\frac{\partial}{\partial\theta}\theta^2 = 2\theta \qquad \frac{\partial}{\partial\theta}(\theta) = 1$$

$$X^T\theta^T \quad Y^T X\theta$$

$$Y_i \cdot Y^T \quad Y^T \cdot Y$$

$$\frac{\partial}{\partial\theta}\left(\overline{J(\theta)}\right) = 0$$

$$\frac{\partial}{\partial\theta}\left(X^T\theta^T X\theta - YX^T\theta^T - Y^T X\theta + \cancel{YY^T}\right) = 0$$

$$\frac{\partial}{\partial\theta}\left(X^T\theta^T X\theta - 2Y^T X\theta\right) = 0$$

$$2X^T X\theta - 2Y^T X = 0 = X^T X\theta = Y^T X$$

$$\frac{\partial}{\partial\theta}\left(X^T\theta^T X\theta\right)$$

$$= \cancel{\theta^T\theta \frac{\partial}{\partial\theta}(X^T X)} +$$

$$X^T X \frac{\partial}{\partial\theta}(\theta^T\theta)$$

$$\theta = (X^T X)^{-1} Y^T X$$

$$X^T X \theta = Y^T X \Rightarrow \theta = \frac{Y^T X}{X^T X} = (X^T X)^{-1} Y^T \cdot X$$

$$\theta_j := \theta_j - \alpha \cdot \frac{\partial J(\theta)}{\partial \theta} \Rightarrow \theta_j$$

Learning Rate Learning Objective.

Coef. intercept

$R^2 \Rightarrow$ Accuracy Measure $\Rightarrow$

$$1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \Rightarrow \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$$

Logistic Regression ⇒

Binary classification ⇒ $h_\theta(x) = \theta_0 + \theta_1 x.$

Sigmoid Function ⇒

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

$$= \frac{e^{x}}{1+e^{x}}$$

$$\cong 0$$

$x = -\infty \cong 0$

$x = \infty \cong 1$

$0 \leq f(x) \leq 1$

$$f(h_\theta(x)) = \frac{h_\theta(x)}{1 + h_\theta(x)}$$

$$0 \le f(h_\theta(x)) \le 1$$

Probability of $x \in (y)$

Predict-Proba

0.91 0.09

0.6

0.2

0 → 1 ⇒

False True

Decision Boundary ⇒

$Pr(x) \geq 0.5 \Rightarrow 1$

$Pr(x) < 0.5 \Rightarrow 0$

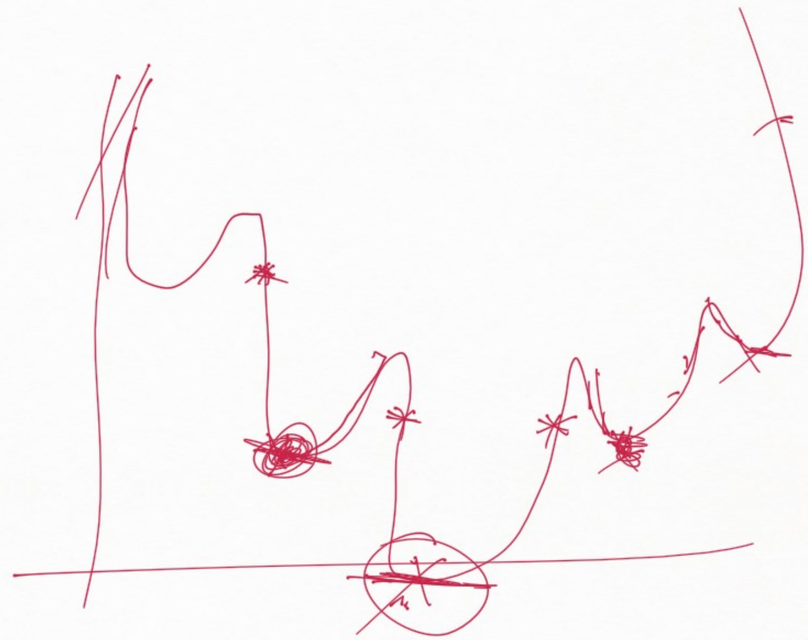
$P(A) + P(B) = 1$

$0.6 \quad 0.4$

logit $\Rightarrow \dfrac{e^{\theta_0 + \theta_1 x}}{1 + e^{(\theta_0 + \theta_1 x)}} \Rightarrow$

R.S.S $\Rightarrow (y_i - \hat{y}_i)^2$

$J(\theta) = (y_i - \hat{y}_i)^2$



$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(h_\theta(x_i)) + (1-y_i)\log(1-h_\theta(x_i))\right]$$

$\log 0 \Rightarrow$

$$\begin{cases} y_i = 0,\ h_\theta(x_i) = 0 & = 0 \\ y_i = 1,\ h_\theta(x_i) = 1 & = 0 \end{cases}$$

$$\Rightarrow \begin{cases} y_i = 0,\ h_\theta(x_i) = 1 \Rightarrow -\frac{\infty}{n} \\ y_i = 1,\ h_\theta(x_i) = 0 \Rightarrow \infty \end{cases} \Bigg\} = \infty$$

$$\text{Odds} \Rightarrow \frac{P}{1-P}$$

$$\text{log-odds} \Rightarrow \log\left(\frac{P}{1-P}\right) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$$