Supervised Learning $\Rightarrow$

Dataset $\Rightarrow$ Features $\Rightarrow$ Input Variable

Labels / targets. $\Rightarrow$ Prediction

Regression

Quantitative

$\in \mathbb{R}$

Price sales estimate.

Classification

Qualitative / Discrete

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

2-classes $\Rightarrow$ True/False $\Rightarrow$ Binary classification

3- " $\Rightarrow$

n- "

$>2 \Rightarrow$ 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 $\Downarrow$ Dependent Variable

θ_0 $\Downarrow$ Intercept

θ_1 $\Downarrow$ Slope

x_1 $\Downarrow$ Independent Variable

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

$\theta_0 . 1 = \theta_0 x_0 = 1$

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n \Rightarrow$ 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 ⇒

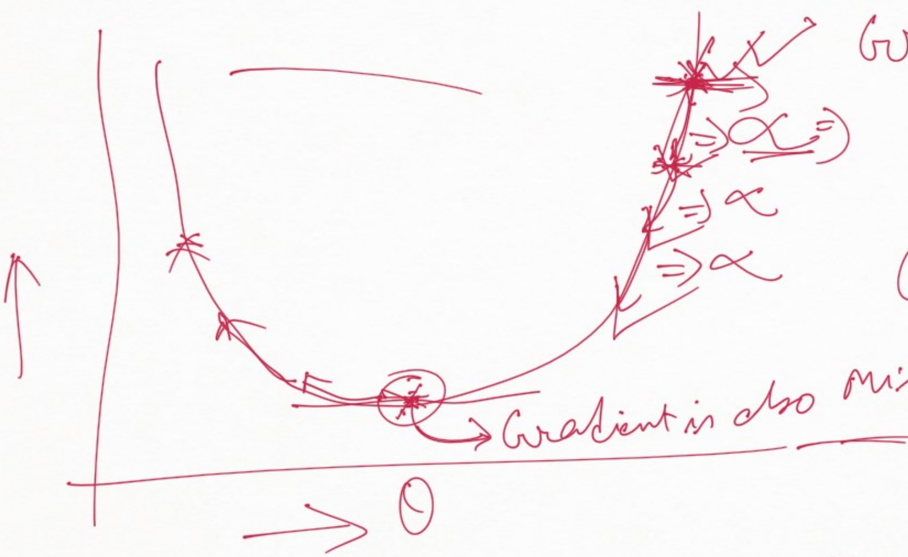
i) Find out optimal $\theta \Rightarrow [\theta_0\ \theta_1 \ldots \theta_n]$

ii) Measure Accuracy

i) Cost Function ⇒ $J(\theta) \Rightarrow$ 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

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

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

$$\frac{\partial}{\partial \theta}\left(J(\theta)\right) = 0 \qquad J(\theta) = \sum_{i=1}^{n} \left(h_\theta(x_i) - y_i\right)^2$$

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

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

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

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

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

$$YX^T\theta^T \qquad Y.Y^T$$
$$Y^T X\theta \qquad Y^T.Y$$

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

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

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

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

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

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

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

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

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 ⇒ $\underline{h_\theta(x)} = \underline{\theta_0 + \theta_1 x}$,

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 $\rightarrow$ 1 $\Rightarrow$

False True

Decision Boundary $\Rightarrow$

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

$Pr(x) < 0.5 \Rightarrow 0$

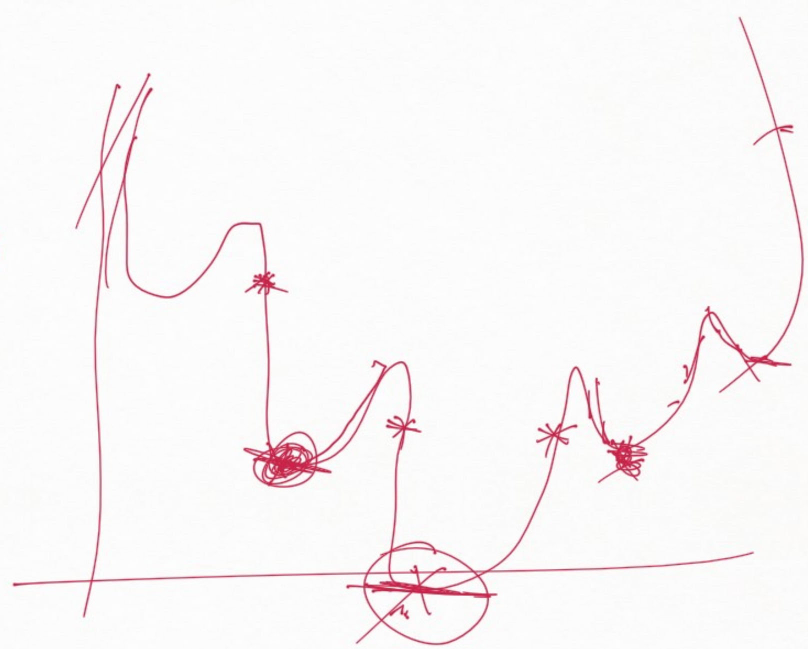
$P(A) + P(B) = 1$

$0.6 \quad 0.4$

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

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

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



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

$\log 0 \Rightarrow$

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

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

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

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

Maximum Livelihood Estimation:

$$0 \leq sig(h_\theta(x)) \leq 1 \Rightarrow \begin{matrix} P(x) \in True \\ P(x) \in False \end{matrix}$$

$$\underline{P(True) + P(False) = 1}$$

$\underline{P}$, $\underline{1-P}$

$$P = P(x) = \frac{e^z}{1+e^z} \quad \text{where } z = \underline{h_\theta(x)} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots \theta_n x_n$$

$$1 - P = 1 - \frac{e^z}{1+e^z} = \frac{1}{1+e^z}$$

$$\frac{P}{1-P} = \frac{e^z}{1+e^z} \times \frac{1+e^z}{1} = e^z = e^{\theta_0 + \theta_1 x_1 + \dots + \theta_n x_n}$$

$$\log_e\left(\frac{p}{1-p}\right) = \log_e\left(e^{z}\right) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

log-Odds $\Rightarrow$ Maximize $\leftrightarrow$ P

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

True	False
0.9	0.1
0.5 $\longleftrightarrow$	0.5
0.3	0.7

$$\Rightarrow \frac{0.9}{0.1} = 9$$

$$\frac{0.5}{0.5} = 1$$

$$\frac{0.3}{0.7} = 3/7$$

$$0 < \text{sig}(h_\theta(x)) < 1$$

Multinomial Log. Reg. $\Rightarrow$

K - Classes. $\Rightarrow$ A, B, C

Baseline $\Rightarrow$ $P(C)$ $\Rightarrow$ Calculate $P(A)$ & $P(B)$ using $P(C)$

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

$\log\left(\frac{P(A)}{P(C)}\right) = \theta_0 + \theta_1 x_1$

$P(A) = P(C) \cdot e^{\theta_{A0} + \theta_{A1} x_1}$

$P(B) = P(C) \cdot e^{\theta_{B0} + \theta_{B1} x_{B1}}$

$P(C)\left(1 + e^{\theta_{A0} + \theta_{A1} x_{A1}} + e^{\theta_{B0} + \theta_{B1} x_1}\right) = 1$

$P(C) = \frac{1}{\{\ \}}$

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

$P(A), P(B), P(C)$

$\underline{K} = \{1, 2, \ldots K-1\}, K$

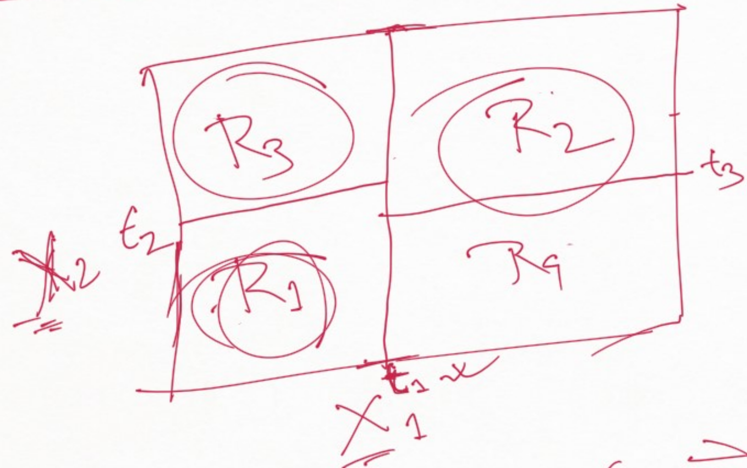
$$\Rightarrow \boxed{P(Y = k \mid X = x)} = \frac{1}{1 + \sum_{i=1}^{K-1} e^{(\theta_{i0} + \theta_{i1} x_1 + \cdots + \theta_{in} x_n)}}$$

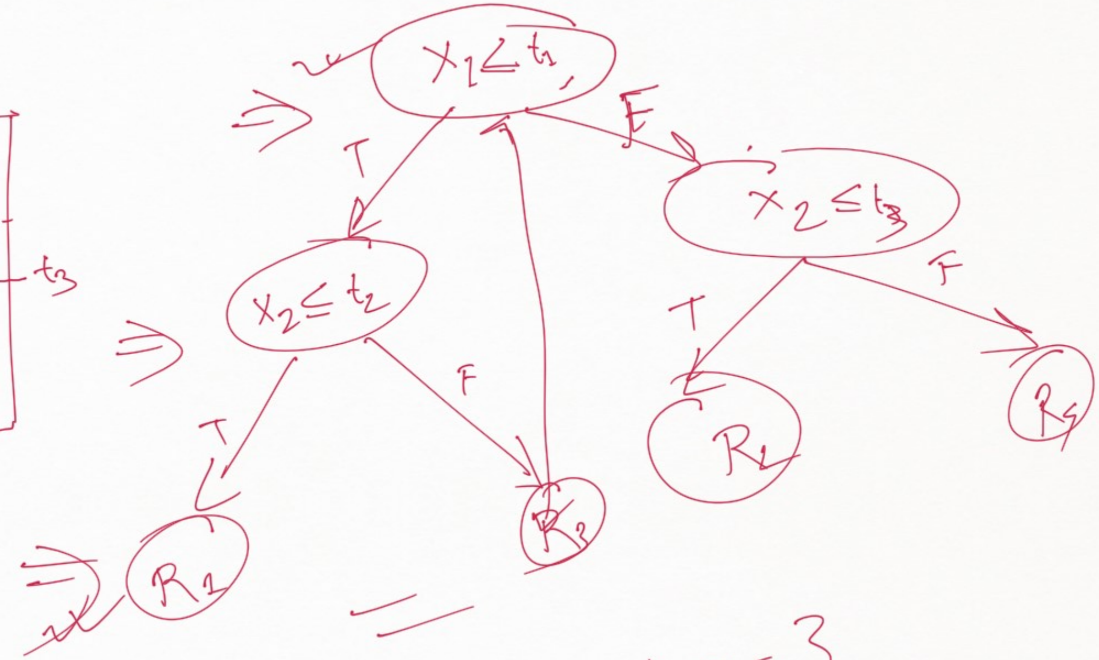
$$P(Y = K \mid X = x) = \frac{1}{1 + \sum_{i=0}^{K-1} e^{(\theta_{i0} + \theta_{i1} x_1 + \cdots + \theta_{in} x_n)}}$$

$$\Rightarrow \frac{P(Y = k \mid X = x)}{P(Y = K \mid X = x)} = e^{\theta_{k0} + \theta_{k1} x_1 + \cdots + \theta_{kn} x_n}$$

$$\log(\ ) = \theta_{k0} + \theta_{k1} x_1 + \cdots + \theta_{kn} x_n$$

Decision Tree





Recursive Binary Division of Input Space

CART

Depth of a tree = 3

Regression $\Rightarrow R_1 \Rightarrow \{ \Rightarrow \} =$ Mean of $\underline{Y}$ in $R_1 \Rightarrow$

$R_2 \Rightarrow$

Cost Function $\Rightarrow$ Squared Error $\Rightarrow \underline{\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2} \Rightarrow$ J Regions are there

$J=2 \Rightarrow R_1, R_2$

$$\underline{\sum_{i \in R_1} (y_i - \hat{y}_{R_1})^2 + \sum_{i \in R_2} (y_i - y_{R_2})^2}$$

Classification $\Rightarrow$

i) Take ratio of inputs in $R_1 \Rightarrow$ Class $1 \Rightarrow 60\% \Rightarrow R_1 = 1$

$R_2 \Rightarrow$ Class $k \Rightarrow 70\% \Rightarrow R_2 = k$

$$Pr(Y = k \mid X \in R_j)$$

ii) Entropy, Gini Impurity

$$\text{Entropy} = -\sum_{l=0}^{K} Pr(Y=l)\log\left(Pr(Y=l)\right)$$

If $K=2$, $\Rightarrow$

$$= -Pr(Y=0)\log\left(Pr(Y=0)\right) - Pr(Y=1)\log\left(Pr(Y=1)\right)$$

$$= -Pr(Y=0)\log\left(Pr(Y=0)\right) - \left(1-Pr(Y=0)\right)\log\left(1-Pr(Y=0)\right)$$

$P(Y=0)+P(Y=1)=1$

$Pr(Y=0)=\frac{1}{2} \Rightarrow$

$Pr(Y=0)=0.5$

$$= -0.5\log(0.5)\times 2 = -\log(0.5)$$

$$= -\log(2^{-1}) = 1$$

Gini Impurity $\Rightarrow$

$$\sum_{l=0}^{K} \Pr(Y=l)\left(1-\Pr(Y=l)\right) = \Pr(Y=0)\left(1-\Pr(Y=0)\right) + \Pr(Y=1)\left(1-\Pr(Y=1)\right)$$

$$= \Pr(Y=0)\left(1-\Pr(Y=0)\right) + \left(1-\Pr(Y=0)\right)\left(1-1+\Pr(Y=0)\right)$$

$$= 2\Pr(Y=0)\left(1-\Pr(Y=0)\right) = 0$$

$K=2 \rightarrow$

$\Pr(Y=0)+\Pr(Y=1)=1$

$\Pr(Y=1)=1-\Pr(Y=0)$

$\Pr(Y=0)=1 \mid 0$

$\Pr(Y=0)=0.5$

$$= 2\times 0.5\times 0.5 = 0.5$$

$\Pr(Y=0)=0.7$

$$= 2\times 0.7\times 0.3 = 0.42$$

Overfitting

Cost Complexity Pruning $\Rightarrow$ Cost function + $\alpha |T|$

$|T| \Rightarrow$ Length of a Tree

Depth " "