Day 2 – Live Machine Learning Algorithm

Agenda

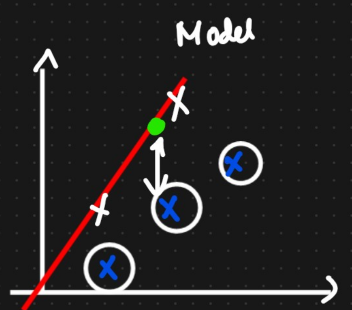
1. Ridge and Lasso Regression
2. Assumption of Linear Regression
3. Logistic Regression
4. Confusion Matrix
5. Practicals Implementation

① Ridge And Lasso Regression

$$\boxed{\text{Cost function} : J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{i}) - y^{(i)}\right)^2}$$

$\theta_0 = 0$

$J(\theta_0,\theta_1) = 0 \downarrow\downarrow\downarrow$



{Training data}

Overfitting ✓

Model performs well → Training data $\boxed{\text{(Low Bias)}}$

Fails to perform well → Test Data ✓ $\boxed{\text{(High Variance)}}$

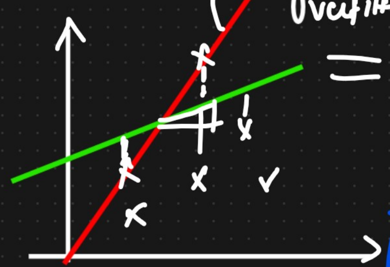
Underfitting { High Bias, High Variance }

{ ① Model Accuracy is bad with Training data
② Model Accuracy is also bad with Test data }

Model 1	Model 2	Model 3
Traing Acc = 90%	Traning Acc = 92%	Traning Acc = 70%
Test Acc = 80%	Test Acc = 91%	Test Acc = 65%
⇓	⇓	⇓
{ Overfitting, Low Bias, High Variance }	{ Generalized Model } { Low Bias, Low Variance }	Underfitting, High Bias, High Variance

slope should not be steep

Overfitting



$\boxed{\lambda = 1}$

$J(\theta_1) = 0$

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

$= (\hat{y}^i - y^{(i)})^2 + \lambda(\text{slope})^2$ ✓

$h_\theta(x) = \hat{y}$ $\theta_1 = 2$ $\theta_0 = 0$

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

$h_\theta(x) = \theta_1 x$ → slope

Ridge (L2 Regularization)

$= 0 + 1(2)^2$

$= 4$ ⇓⇓ ⇓⇓

iterations {Hyperparameter}

R^2, adjusted R^2

$\theta_1 \Rightarrow$

$= (\hat{y}^{(i)} - y^{(i)})^2 + \lambda(\text{slope})^2$ $\lambda \rightarrow$ Hyperparameter ✓

⇓

Convergence

{

}

{ Prevent Overfitting }

(Small value) $+ 1(1.5)^2$

$=$ (Small value) $+ 2.25$

≈ 3 ⇓⇓

feature selection

Lasso (L1 Regularization)

$= (\hat{y} - y)^2 + \lambda \,|\text{slope}|$

$|\theta_0 + \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5 \cdots + \theta_n|$

$h_\theta(x) = \hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \cdots + \theta_n x_n$

① Preventing Overfitting
② Feature Selection } → { L1 Regularization } { L2 Regularization }

⇓

Performance metrics

⇓.

{ λ → Cross validation }

Ridge Regression (L2 Norm)

$$\text{Cost function} = (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda(\text{slope})^2$$

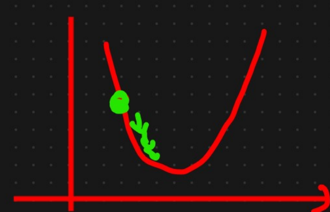
④ **Purpose**: Preventing Overfitting

$|\theta_0 + \theta_1 + \theta_2 + \theta_3 \cdots + \theta_n|$

Lasso Regression (L1 Reg)

$$\text{Cost function} = (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda|\text{slope}|$$

Purpose:
1) Prevent Overfitting
② Feature Selection



Assumption of Linear Regression

① Normal / Gaussian Distribution → Model will get trained well

✓② {Standardization {Scaling data} → Z-score $\mu=0, \sigma=1$}

③ Linearity

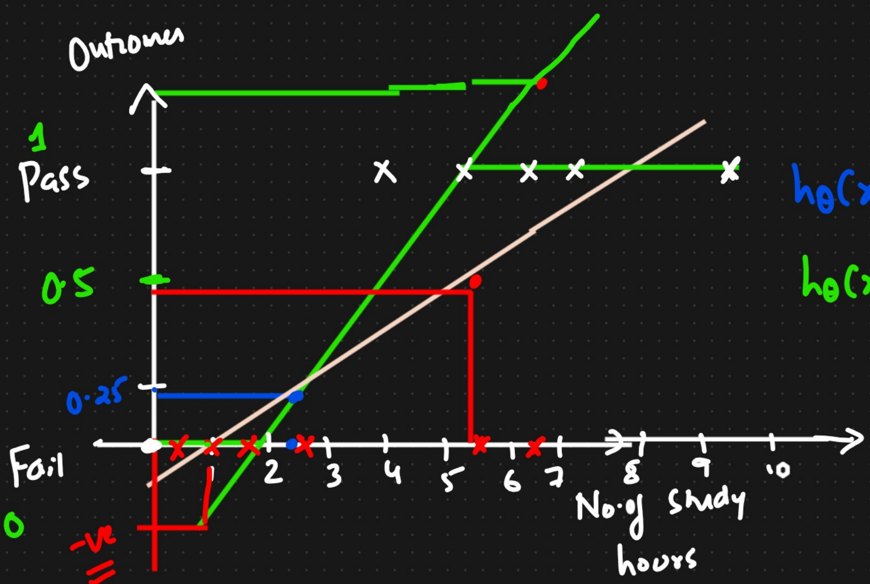
X_3 $\boxed{X_1 \; \cancel{X_2}}$ ↗95% $\boxed{Y}$

$\boxed{\text{Variation Inflation factor?}}$

④ Multi Collinearity

Logistic Regression (classification) → Binary classification

No. of study	No. of play	P/F
—	—	P
—	—	F



Linear Regression ?? ✓

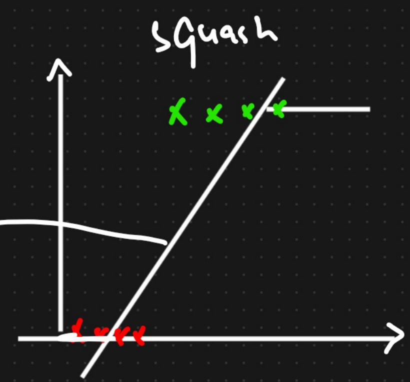
$h_\theta(x) < 0.5 \Rightarrow 0 \rightarrow$ Fail

$h_\theta(x) \geq 0.5 \Rightarrow 1 \rightarrow$ PASS

{ Sigmoid function }

$\boxed{0 \text{ to } 1}$ ✓

Decision Boundary Logistic Regression



$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 - \cdots + \theta_n x_n$$

$$\boxed{h_\theta(x) = \theta^T x}$$

$$h_\theta(x) = \boxed{\theta_0 + \theta_1 x_1}$$

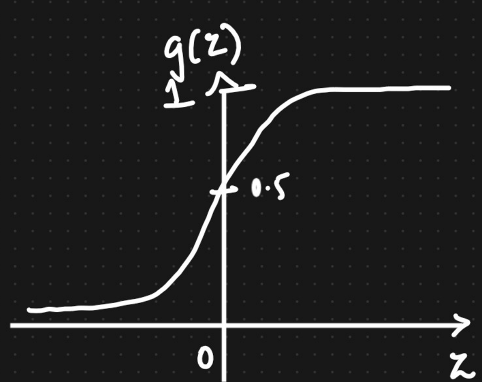
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1)$$

Let $z = \theta_0 + \theta_1 x$

$$h_\theta(x) = g(z)$$

$$h_\theta(x) = \frac{1}{1+e^{-z}}$$

Sigmoid or Logistic function



$$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}}$$

$g(z) \geq 0.5$ when $z \geq 0$ } ✓.

Training Set

$$\{(x^1,y^1),(x^2,y^2),(x^3,y^3) \ldots (x^n,y^n)\}$$

$$y \in \{0,1\} \rightarrow 2 \text{ o/p}$$

$$h_\theta(z) = \frac{1}{1+e^{-z}} \qquad \boxed{z = \theta_1 x}$$

Change parameter θ_1 ?

Cost function

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

Linear Regression $J(\theta_1) = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{2} \overbrace{(h_\theta(x^i) - y^i)^2}^{\text{cost function}}$

Logistic Regression $\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_1 x)}}}$

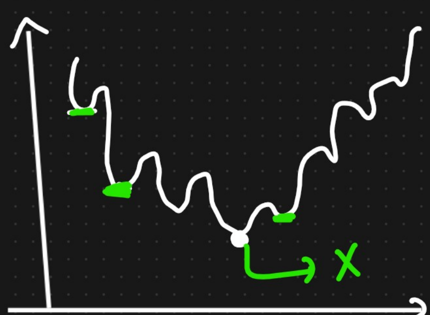
Logistic Regr Cost function $= \frac{1}{2}(h_\theta(x^{(i)}) - y^{(i)})^2$

We cannot use this cost function for logistic

$$h_\theta(x) = \frac{1}{1+e^{-(\theta_1 x)}}$$

Gradient Descent

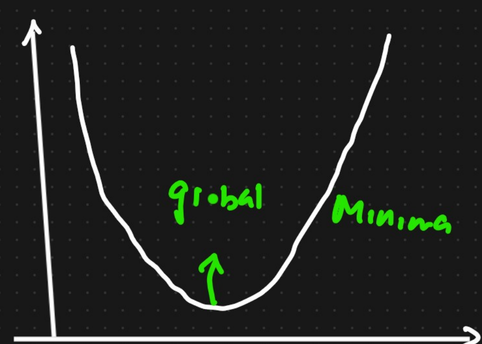
Non convex function



Local Minima Problem

Gradient Descent

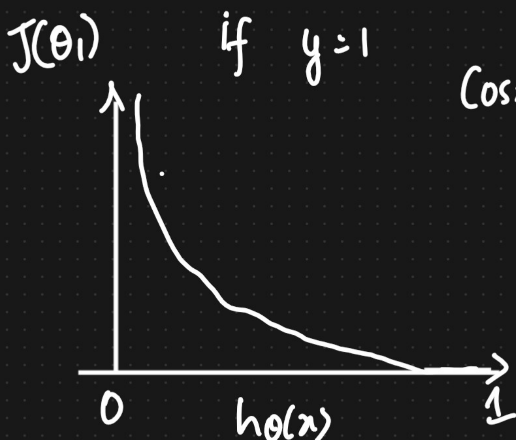
Convex function



Logistic Regression Cost function

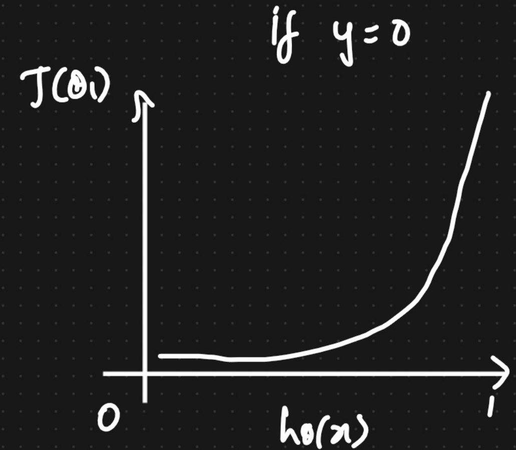
$$h_\theta(x) = \frac{1}{1+e^{-(\theta_1 x)}}$$

$$J(\theta_1) = \begin{cases} -\log(h_\theta(x^i)) & y=1 \\ -\log(1-h_\theta(x^i)) & y=0 \end{cases}$$



Cost $=0$ if $y=1$, $h_\theta(x)=1$

Classification



$$\text{Cost}(h_\theta(x^i), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0 \end{cases}$$

$$\boxed{\text{Cost}(h_\theta(x^i), y) = -y\log(h_\theta(x^i)) - (1-y)\log(1-h_\theta(x))}$$

$\Downarrow$ cost function.

if $y=1$

$$\text{Cost}(h_\theta(x^i), y)) = -\log(h_\theta(x^i))$$

if $y=0$

$$\text{Cost}(h_\theta(x^i), y) = -\log(1-h_\theta(x^i))$$

$$J(\theta_1) = -\frac{1}{2m} \sum_{i=1}^{m} \left[\left(y^i \log(h_\theta(x^i)) + (1-y^i) \log(1-h_\theta(x^i)) \right) \right]$$

$\downarrow$

cost

$$h_\theta(x^i) = \frac{1}{1+e^{-\theta_1 x}}$$

repeat until convergence

$\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\theta_j} \left(J(\theta_j) \right)$$

$\}$

Performance Metrics {Classification problem}

x_1	x_2	y (Actual)	$\hat{y}$ (o/p)
–	–	0	1
–	–	1	1
–	–	0	0
–	–	1	1
–	–	1	1
–	–	0	1
–	–	1	0

{Actual}

Pred	1	0
1	3	2
0	1	1

Actual

Predicted	1	0
1	TP	FP$\downarrow$
0	FN$\downarrow$	TN

$\rightarrow$ Confusion matrix

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$= \frac{3+1}{3+2+1+1} = \frac{4}{7}$$

$$= 0.57 = 57\%$$

② $0 \rightarrow 900$, $1 \rightarrow 100$ } Imbalance DATASET

$0 \rightarrow 600$, $1 \rightarrow 400$ } Balanced Data

$0 : 900$, $1 : 100$ } Biased

$$\left\{ \text{Model} \rightarrow 0 = \frac{900}{1000} = 90\% \right\}$$

. TPR, Sensitivity

① Precision

$$\frac{TP}{TP+FP}$$

② Recall

$$\frac{TP}{TP+FN}$$

③ F-Score.

{Tom Stock market is going to crash}

{Spam classification} → Precision

{HAS CANCER OR NOT} → Recall

		Actual 1	Actual 0
Pred	1	TP	FP
	0	FN	TN

$$\text{F-Beta} = (1+\beta^2)\,\frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}$$

$\beta = 1$

$$= (1+1)\,\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-Score

$$= \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Harmonic Mean

$$= \frac{2xy}{x+y}$$

$\beta = 0.5$

$$(1+(0.5)^2)\,\frac{P \times R}{(0.25)P + R}$$

F0.5 Score

$\beta = 2$ FN >> FP.

F2 Score