Machine Learning Algorithms

1. Ridge and Lasso Regression ← {Overfitting, Bias, Variance, Underfitting}
2. Elastic Net Regression
3. Logistic Regression

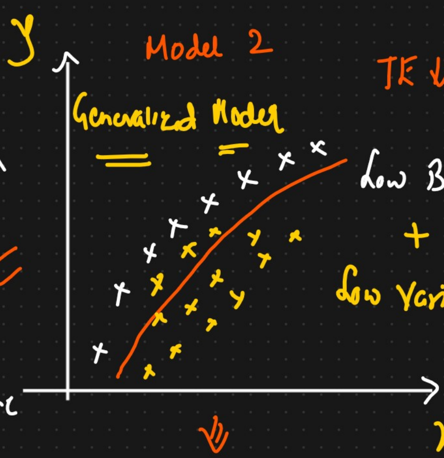
{ Bias = Training Data
Variance = Test Data }

Training Data
Test Data ↓↓↓
Acc

Model 1 — Underfitting

{ Training Error ↑↑ }
High Bias ✓
+
Low Variance
{ or
High Variance }

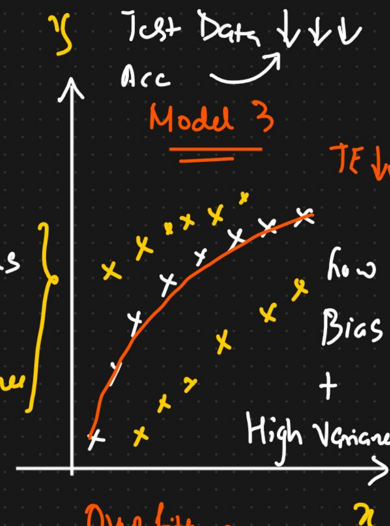
Model 2 — Generalized Model



TE ↓↓
Low Bias
+
Low Variance

⇓
Generalized Model

Model 3



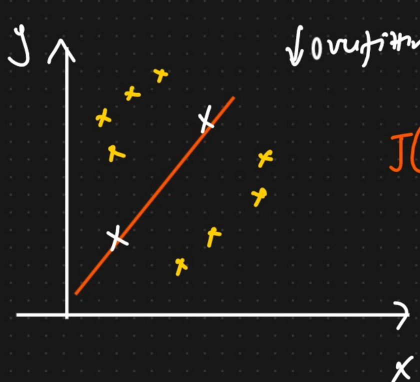
TE ↓↓
Low Bias
+
High Variance

Overfitting

{ Training Data → Train our Model ✓ }
{ Test Data → Check the performance of the Model ✓
↳ Validation Data }

② Ridge and Lasso Regression

Train Error ↓ + Test Error ↑↑
{Low Bias + High Variance}

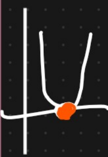


↓overfitting

Cost function

$$J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{i}) - y^{(i)}\right)^2 = 0$$

Ridge Regression (L2 Regularization) $\boxed{\lambda = 1}$

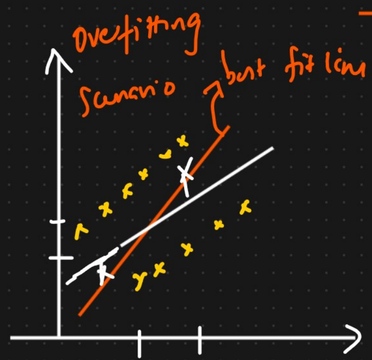


$$J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + \lambda\,(\text{slope})^2$$

Hyperparameter

Penalty

Slope



$= 0 + (1)(1.5)^2$

$= 2.25$ ↓↓↓↓

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

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

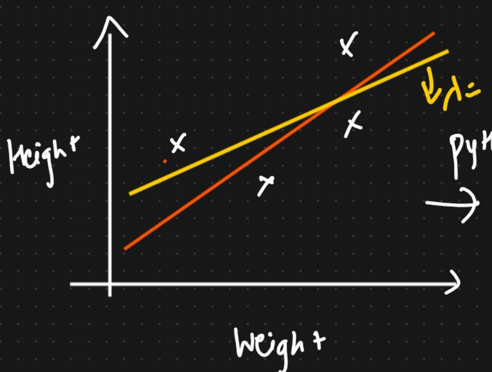
$(\theta_1^2 + \theta_2^2 + \theta_3^2 + \theta_4^2)$

$= \text{Small value} + (1)(0.9)^2$

≈ 0.81 ↓↓

Relationship between λ and $(\text{Slope})^2$

λ ↑↑



python → $\boxed{\lambda}$ → Hyperparameter Tuning } Cost

$$J(\theta_0,\theta_1) = \underbrace{\frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2}_{} + \overbrace{\lambda\,(\text{Slope})^2}^{\text{Penalty}}$$

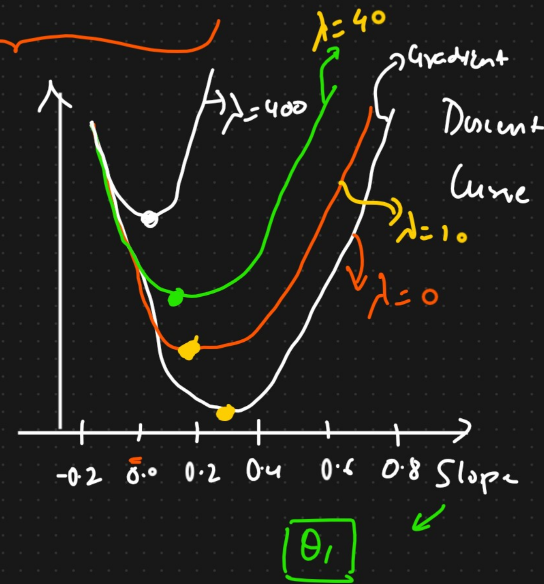
λ = Different λ values

Will get intilized

⇓

Hyperparameter

$\lambda = 10$



$\boxed{\lambda \uparrow\uparrow}$ $\boxed{\text{Slope} \downarrow\downarrow}$

Global Minima Is Shifting

λ ↑↑ Slope ↓↓

⇓ ⇓ ⇓⇓

{Reduce Overfitting}

$h_\theta(x) = \theta_0 + \theta_1 x$ → neglecting this feature.

$\theta_1 \to 0$

Reduce Overfitting

Lasso Regression (L_1 Regularization) ⇒ Feature Selection

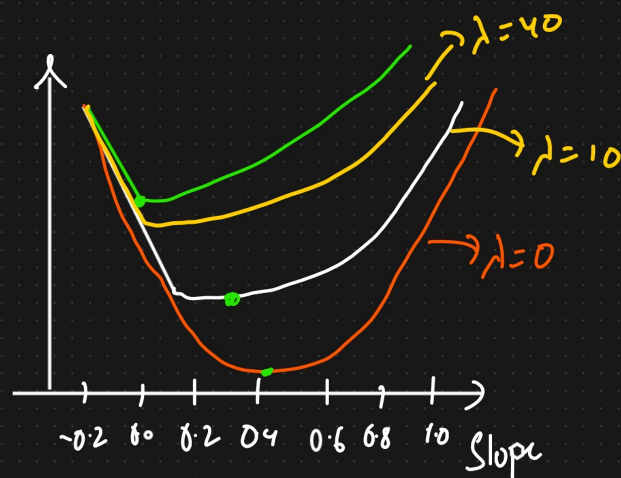
$$\text{Cost function } J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{i}) - y^{(i)} \right)^2 + \lambda |\text{Slope}|$$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots + \theta_n x_n$$

$\theta_3 \to 0.13 \to 0$

$\lambda \uparrow\uparrow$ Slope $\downarrow\downarrow$

{ 1 unit x, 0.13 unit y }



⇒ feature selection }

$\lambda \uparrow\uparrow$ Slope $\downarrow\downarrow$

Not at all correlated ≈ 0 → feature neglected ⇒ feature selection.

Final Conclusion

{ Ridge = Reduce Overfitting
Lasso = Feature Selection } ⇒ Overall

Elastic Net Regression ✓

Reducing overfitting ⇑ Feature Selection ⇑

$$\boxed{J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y^{(i)} \right)^2 + \lambda_1 (\text{Slope})^2 + \lambda_2 |\text{Slope}|}$$

Multicollinearity

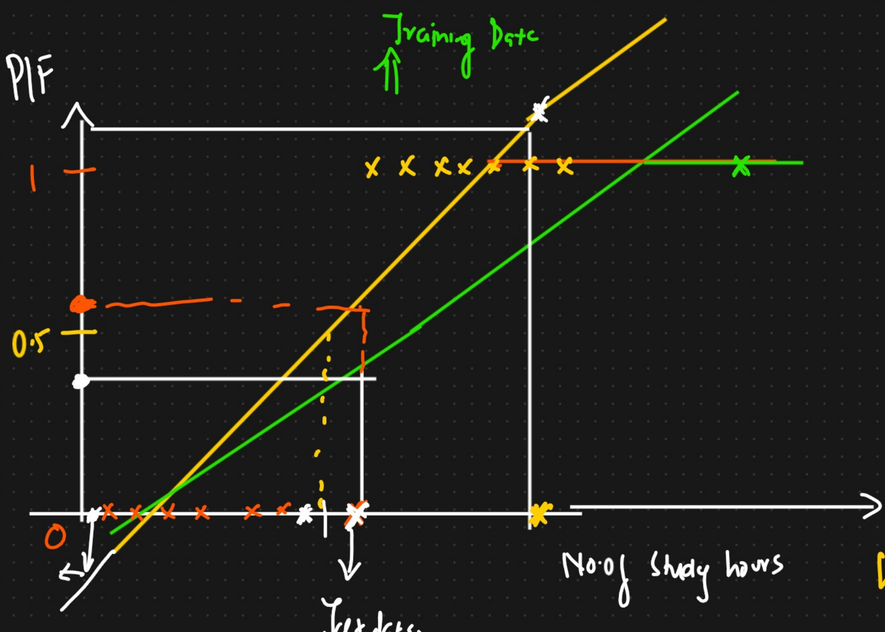
$\cancel{f_1} \longleftrightarrow f_2 \quad f_3 \qquad O/P$

0.95%

-0.95

600 features $\Rightarrow$ Lasso

② Logistic Regression (Classification) $\rightarrow$ { Binary classification }



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

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$

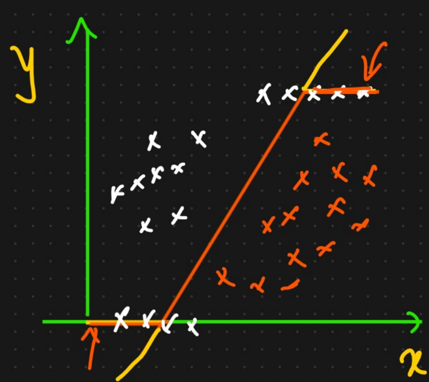
$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1}$$

$\Downarrow$

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

$\Downarrow$

Let $z = \theta_0 + \theta_1 x_1$



$h_\theta(x) = g(z)$ $\rightarrow$ function on z

$\Downarrow$

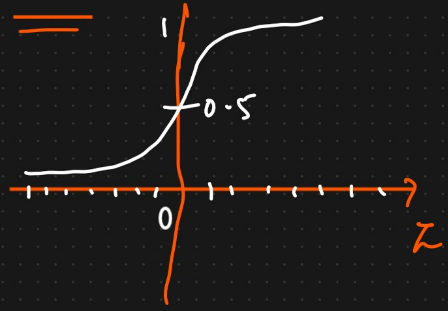
{ Activation function }

Sigmoid Activation fn

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

O{ 0 to 1 }

exponential



$$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x_1)}}}$$

$\Downarrow$

Best fit line + Squashing

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

Training Set

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

$y=\{0,1\} \rightarrow$ 2 Output $\rightarrow$ Binary classification

$$h_\theta(z) = \frac{1}{1+e^{-z}} \qquad z = \theta_0 + \theta_1 x_1 \qquad \theta_0 = 0 \qquad \text{intercept} = 0$$

Aim : Change $\theta_1 \rightarrow$ It classifies points.

Costfunction

Logistic Regression

$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\Big(\underbrace{h_\theta(x)^i}_{\hat{y}} - y^{(i)}\Big)^2 \quad \leftarrow \text{MSE}$$

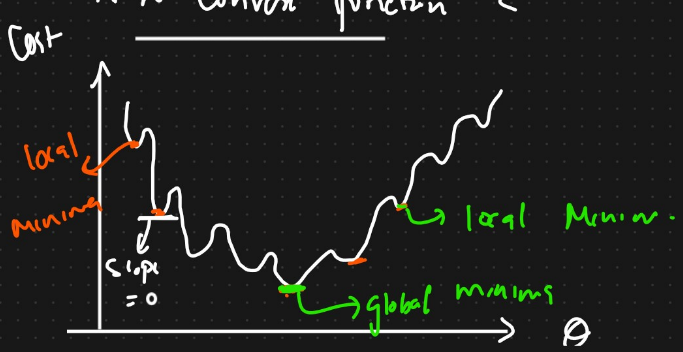
$h_\theta(x) = \theta_1 x_1$

$\Downarrow$

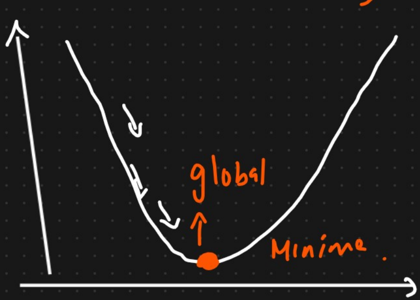
Gradient Descent

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

Non Convex function



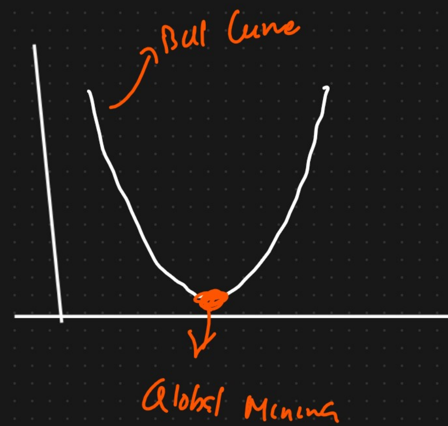
{ Convex function }



Logistic Regression Cost function (Log Loss)

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

$h_\theta(x) \to \hat{y}$



$y=1$

Cost function

$$J(\theta_0,\theta_1) = -y\log(h_\theta(x^i)) - (1-y)\log(1-h_\theta(x))$$ {log loss Cost function}

Cost function for Logistic Regression

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

$$\boxed{h_\theta(x^i) = \frac{1}{1+e^{-(\theta_1 x_i)}}} \Rightarrow \text{hypothesis.}$$

θ_1

repeat until Convergence

{

$$\theta_j := \theta_j - \alpha\frac{\partial}{\partial\theta_j}(J(\theta_1)).$$

}

Performance Metrics (Binary Classification).

x_1	x_2	Act y	predi $\hat{y}$
–	–	0	1
–	–	1	1
–	–	0	0
–	–	1	1
–	–	1	1
		0	1
		1	0

{Actual}

{predicted}	1	0
1	3	2
0	1	1

⇒ Confusion matrix

Predicted \ Actual	1	0
1	↑TP	FP↓
0	FN↓	TN↑

TP & TN Are correct predictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} = \frac{3+1}{3+2+1+1} = \frac{4}{7} = 57\%$$

Training
1000 datapoint points

0 → 900 datapoint
1 → 100 datapoint } Imbalanced Dataset

0 → 600 datapoints
1 → 400 datapoints } Balanced

Not Spam

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

→ FP ↓↓
→ FN ↓

① Precision

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

② Recall (TPR).

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

③ F-Score

Spam classification {FP↓↓} → Precision

HAS CANCER OR NOT {FN↓↓} → Recall

Company ⇒ FP

People ⇒ FN

Tomorrow Stock market is going to crash $\left\{ \begin{array}{l} \text{Both FP} \downarrow\downarrow \\ \text{\& FN} \downarrow\downarrow \end{array} \right\}$.

F-Beta Score : $(1+\beta^2) \dfrac{\text{Precision} * \text{Recall}}{\beta^2 * [\text{Precision} + \text{Recall}]}$

$\boxed{\text{FP \& FN}}$ ✓

$\boxed{\beta = 1}$ $= (1+1) \dfrac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ } Harmonic Mean

$\beta \downarrow\downarrow$

$\boxed{\text{FP} >> \text{FN}}$

$\beta = 0.5$ $= (1+0.25) \dfrac{P \times R}{(0.25)[P+R]}$

$\boxed{\text{FN} >> \text{FP}}$

$\boxed{\beta = 2}$ $= \boxed{(1+(2)^2) \dfrac{PR}{(4)[P+R]}}$.