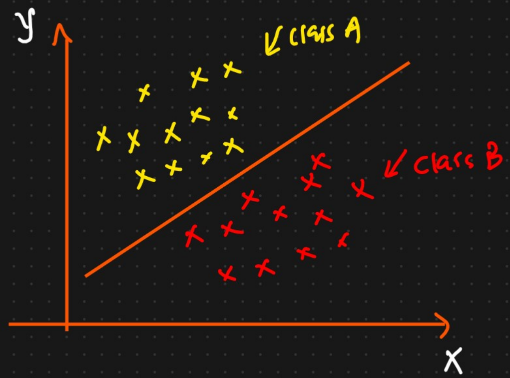
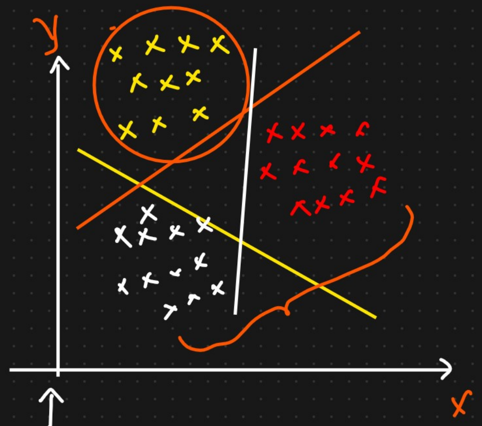
Logistic Regression (One Versus Rest)



① Binary Classification

One Versus One



Multiclass Classification {Logistic Regression}

→ M_1 → Binary classification
→ M_2 → Binary classification
→ M_3 → Binary classification

One Versus Rest (OVR) → Logistic Regression

O_1, O_2, O_3

f_1	f_2	f_3	O/p	$\boxed{O_1}$	O_2	$\boxed{O_3}$
–	–	–	O_1	1	0	0
–	–	–	O_2	0	1	0
–	–	–	O_3	0	0	1
–	–	–	O_1	1	0	0
–	–	–	O_3	0	0	1
–	–	–	O_2	0	1	0

Model

M_1 ← I/p $\{f_1, f_2, f_3\}$ O/p $\{O_1\}$
M_2 ← I/p $\{f_1, f_2, f_3\}$ O/p $\{O_2\}$
M_3 ← I/p $\{f_1, f_2, f_3\}$ O/p $\{O_3\}$

0.55 → O/P = $\boxed{O_3}$ → Category 3

New Test Data → M_1 → 0.25 ✓
→ M_2 → 0.20 ✓
→ M_3 → 0.55 ✓

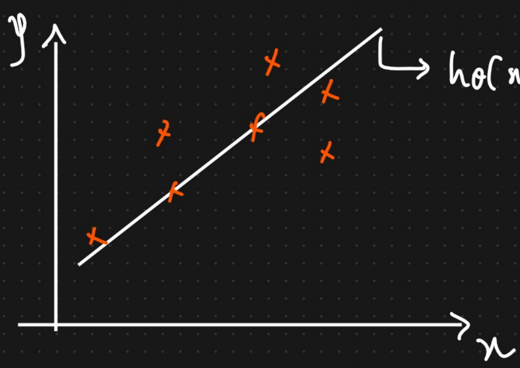
[0.25, 0.20, $\boxed{0.55}$]

M1 M2 $\boxed{M_3}$

O/p

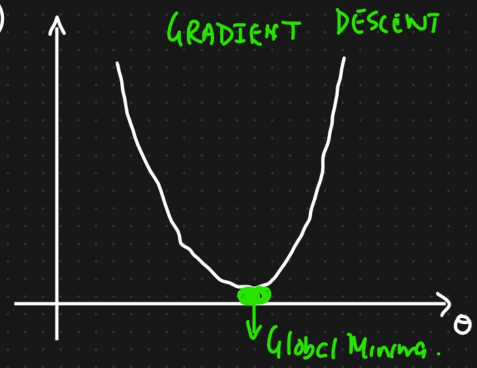
Ridge Regression, Lasso Regression, Elasticnet Regression

Linear Regression



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

Independent features



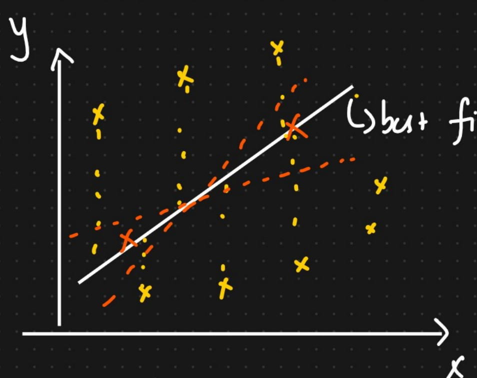
$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$$

$\Downarrow$

Mean Squared Error

① Ridge Regression (L2 Regularization) → Reduce Overfitting

Overfitting



↳ best fit line

Traindata → Acc ↑ → Low Bias

Test data → Acc ↓ → High Variance

$\lambda = 0$

$\lambda = 1$

$\lambda = 10$

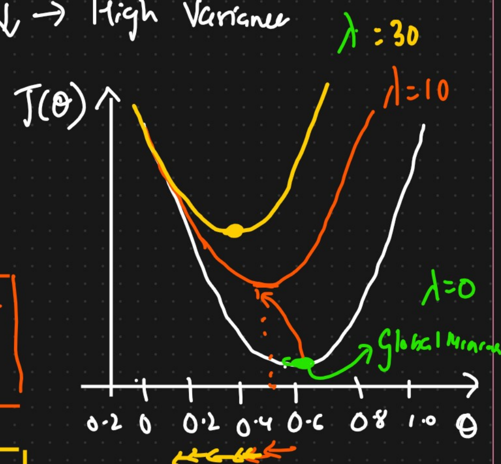
Hyperparameter

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

$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \boxed{\lambda \sum_{i=1}^{n} (\text{slope})^2}$$

$$= \quad 0 \quad + \quad (1)\left[(\theta_1)^2\right]$$

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



λ = Hyperparameter

$\boxed{\lambda = 1}$

> 0

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

0.24

$\Downarrow$

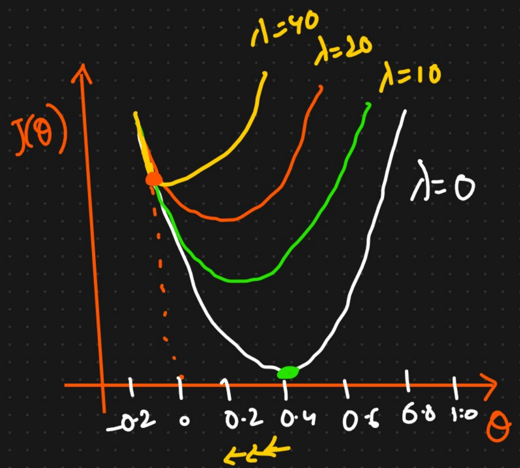
$= 0.34 + \underline{0.52} x_1 + \underline{0.48} x_2 + 0.24 x_3$

$\Downarrow \quad \Downarrow \quad \Downarrow\Downarrow$

$$= 0.34 + 0.40x_1 + 0.38x_2 + \boxed{0.14x_3}$$

② Lasso Regression (L_1 Regularization) → Feature Selection

$$\text{Cost fn} : \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right) + \boxed{\lambda \sum_{i=1}^{n} |\text{slope}|}$$



$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

$$h_\theta(x) = 0.52 + 0.65x_1 + 0.72x_2 + 0.34x_3 + \boxed{0.12}x_4$$

⇓

Lasso Regression

$$= 0.52 + 0.51x_1 + 0.60x_2 + 0.14x_3 + \boxed{\cancel{0 x_4}} \nearrow 0$$

③ ElasticNet Regression
→ ① Reduce Overfitting
→ ② Feature Selection

$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \boxed{\lambda_1 \sum_{i=1}^{m} (\text{slope})^2} + \boxed{\lambda_2 \sum_{i=1}^{m} |\text{slope}|}$$

⇓ Reduce Overfitting

⇓ Feature Selection

{ Hyperparameter Tuning the Linear Regression }

Types of Cross Validation

1000 records

DATASET

70% → TRAINING

30% → TEST ← To check the performance of Model

① Training our model → TRAINING

② Hyperparameter Tuning

TRAINING → TRAIN, VALIDATION

Cross Validation

VALIDATION ← ⇒ Random_state

85%   92%   78%

① Leave One Out CV (LOOCV)

② Leave P out CV

| P=10   P=20   P=30 |

TRAINING → 500 Records ↑↑ Complexity of Training Model

Exp 1: | V | TRAINING | → Acc 1

Validation

Exp 2: | TR | V | TRAINING | → Acc 2

⋮

Exp 500: | TRAINING | V | → Acc_{500}

① Overfitting → TRAINING ↑↑ Acc → New Test Data → Acc ↓↓

Validation Acc ↓↓

③ K Fold CV

K=5 n=500

Binary Classification

1 or 0's

Test size = $\frac{500}{5}$ = 100

CV=1 Exp 1 TRAINING → Acc 1

Validation

Exp 2 → Acc 2

Avg Accuracies

TRAINING Validation

Exp 5 → Acc 5

④ Stratified K Fold CV

K=5

60 1's
40 0's

60:40

Exp 1

TRAINING

Validation

TRAINING

Test size = $\frac{500}{5}$ = 100

⇓

Validation Data

⑤ Time Series CV

Reviews

Product Sentiment Analysis

Time

JAN → DEC

TRAINING

Validation

DAY 1 DAY 2 DAY 3 DAY 4 . - - - - DAY N

Time Series Application

Performance Metrics, Accuracy, Precision, Recall And F-Beta

Topics to be covered

1. Confusion matrix
2. Accuracy
3. Precision
4. Recall
5. F-Beta Score

Logistic Regression

R squared

Adjusted R squared

1. Confusion Matrix

Actual Values

Predicted Values	1	0
1	3	2
0	1	1

Dataset

O/P; $\hat{y}$ = pred by model

x_1	x_2	y	$\hat{y}$
-	-	0	1
-	-	1	1
-	-	0	0
-	-	1	1
-	-	1	1
-	-	0	1
-	-	1	0
-	-		

Actual

Predicted	1	0
1	TP	FP
0	FN	TN

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

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

$$= \frac{4}{7}$$

Dataset — Binary classification

↳ 1000 datapoints → 900 → 1, 100 → 0 } Imbalanced Dataset

⇓

90% Accuracy

④ **Precision** $= \dfrac{TP}{TP + FP\downarrow}$ } Out of all the actual value how many are correctly predicted

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

② **Recall** $= \dfrac{TP}{TP + FN\downarrow}$ $\Rightarrow$ Out of all the predicted value how many are correctly predicted

Usecase ①

Spam classification

Mail → Spam
Model → Spam } Good

Mail → Not Spam
Model → Spam } Blunder

		Spam 1	Not Spam 0 (Actual)
Spam	1	TP	FP ✓ $\Downarrow$
Not Spam	0	FN	TN

$\Downarrow$

$$\text{Precision} = \frac{TP}{TP + FP}$$

Usecase 2

To predict whether person has diabetes or not

✓ Truth → diabetes
✓ Model → Doesn't diabetes } Blunder

	Diabetes	No Diabetes
Diab	TP	FP
No Diab	FN	TN

↓ $\Downarrow$
usecase of diseases

$$\text{Recall} = \frac{TP}{TP + \boxed{FN}\Downarrow}$$

Truth → diabetes
Model → " } good

Truth $\longrightarrow$ Not diabetes
Model $\longrightarrow$ Diabetes $\Bigg\} \Rightarrow$ 2nd opinion $\Downarrow$ check

Assignment

* Tommorow the stock market will crash or not

Reducing FP$\downarrow$ or FN$\downarrow$

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

① If FP & FN are both important

$\beta = 1$

$$F1 \text{ Score} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \Bigg\} \Rightarrow \text{Harmonic Mean}$$

② If FP is more important FN

$\beta = 0.5$

$$F\,0.5 \text{ Score} = (1+0.25)\frac{P * R}{P + R}$$

③ If FN >> FP

$\beta = 2$

$$\boxed{F\,2 \text{ Score} = (1+4)\frac{P \times R}{P + R}}$$