Regression:
L.R.

Classification:
- Log. R
- KNN
- NB
- DT
- SVM

} Unit 3

Ensemble

0
0
1
0
1

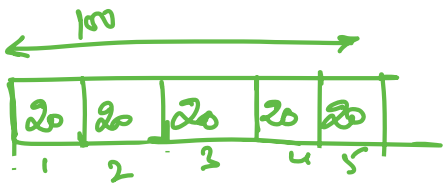
→ 0

[Training (100)] [Test (50) hidden]

Train (80) — Validation (20)

Log. Regression — Acc.: 075%.
NB — Acc.: 80%.
SVM — Acc.: 85%.

k fold cross validation

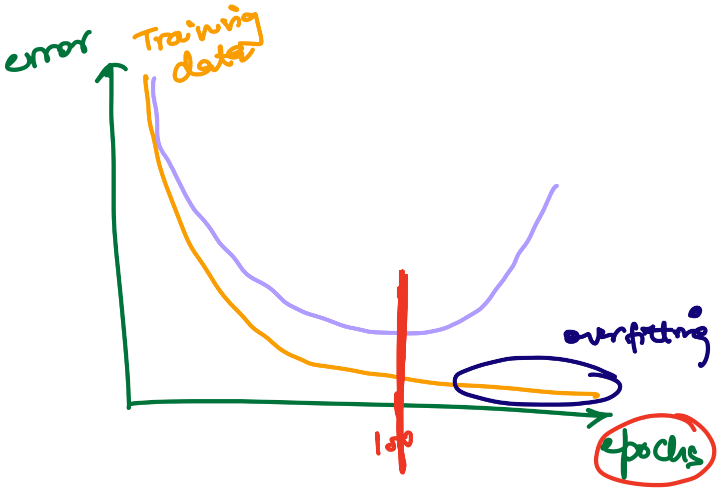


LR

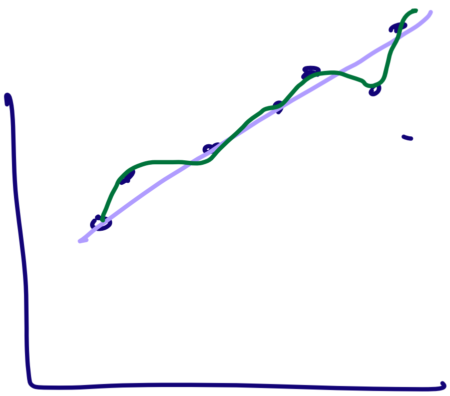
5 : 1st Iter. 1234 Train — 5th valid
2nd Iter. 1235 Train — 4th valid — 70
3rd Iter. 1245 Train — 3rd valid — 70
4th Iter. 1345 Train — 2nd valid — 40
5th Iter. 2345 Train — 1st valid — 30

80

Avg.



↓
hyperparameters



VS

Random w

error (MSE) loss

$w = w - \eta \frac{\partial L}{\partial w}$ → error vs

Parameters → w_1, w_2

Hyperparameters → η

↓ epochs, knn (K), svm (C)

GRID SEARCH

η: 0.1, 0.01, 0.05, 0.001

best result pick
↓
candidate set

Metrics

LR → R2 Score (Regression)

Classification?

Actual labels (True labels)	Logistic Regression Predict	DT Predict	Classification Y, N 1 0
1	1	1	
0	1	1	
0	0	0	
0	0	0	
1	1	1	
1	1	1	
0	1	0	
0	0	0	

$\frac{9}{10} = 90\%$

0	0	0
1	1	1

Accuracy = $\frac{\#\text{correct predictions}}{\#\text{total}}$

$\frac{8}{10} = 0.8 = 80\%$.

Actual	Predicted
0	0
0	0
0	0
2	2
0	0
2	2
0	2
2	1
1	1
1	0

Accuracy = $\frac{7}{10} = 70\%$.

80% → 20% incorrect

Actual	Predicted
0	1
1	0

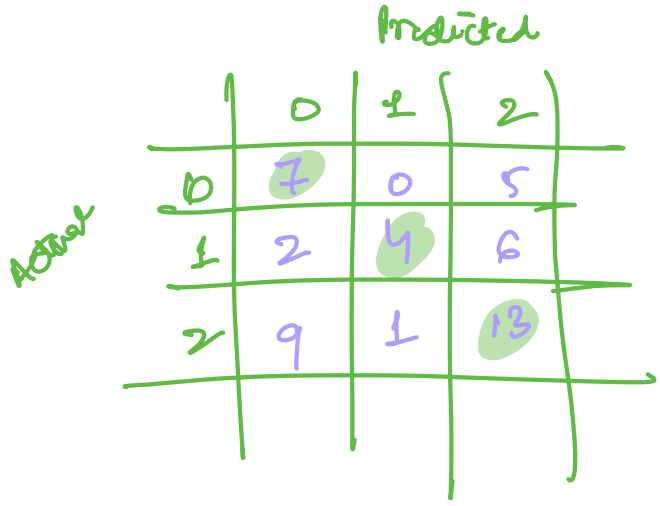
} Accuracy table

Confusion Matrix

Actual \ Prediction	0	1
0	True Negative	False Positive
1	False Negative	True Positive

0: Not Placed
1: Placed

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



$$\text{Acc} = \frac{7+4+13}{7+0+5+2+4+6+9+1+13}$$

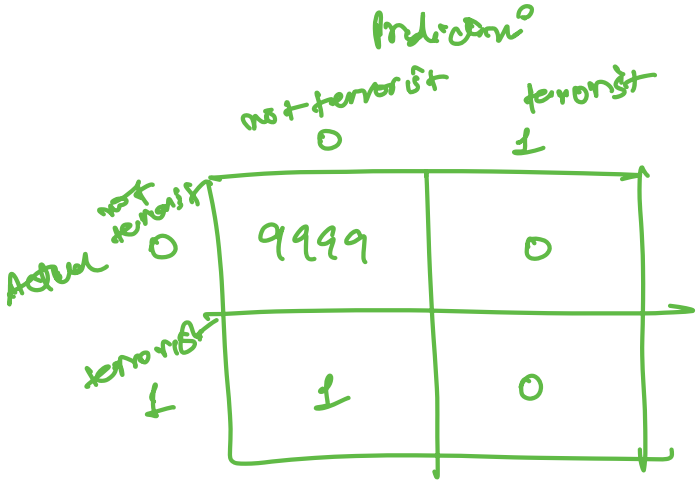
Imbalanced Data

Not terrorist

10,000

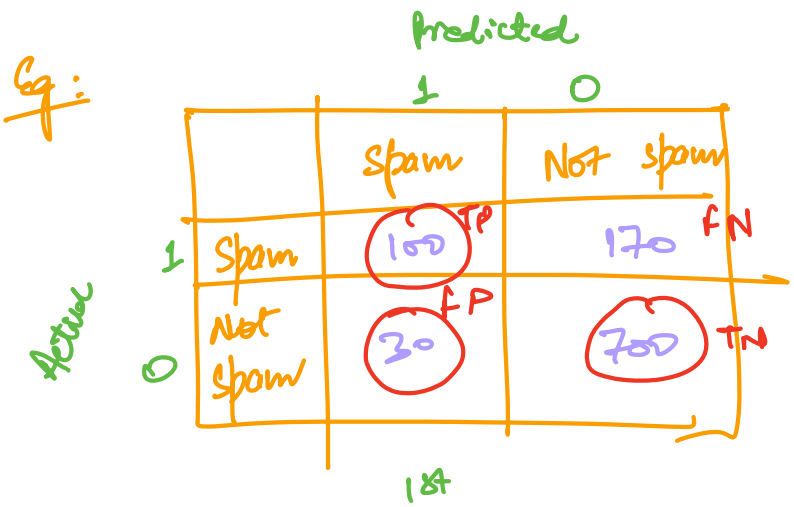
9999 normal

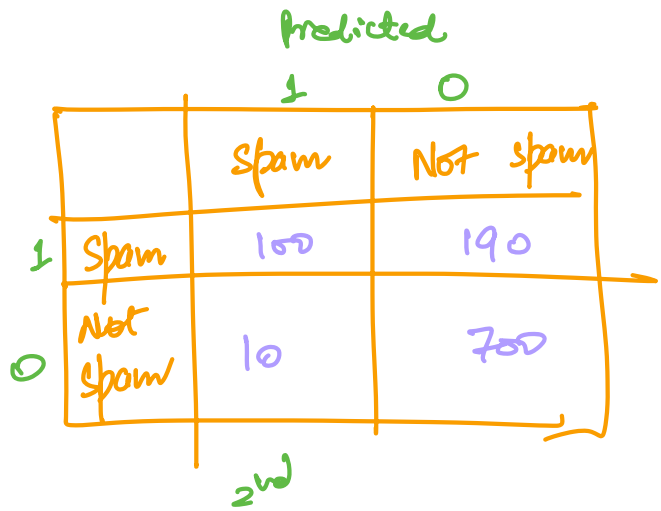
1 terrorist



$$\frac{9999}{9999+1} = 99.99\% \text{ accuracy}$$

Eg:





1st: Valid mail → Spam = 30

2nd: 10

1st: Spam → Inbox: 170

2nd: : 190

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

what proportion of predicted is truly correct.