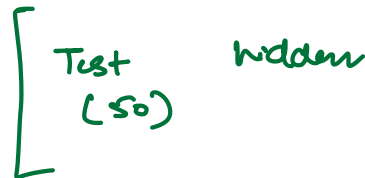Regression:
L·R·

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

Unit 3   Ensemble

0
0
1
0
1
} 0

Training (100)

Test (50)   hidden

Train (80)
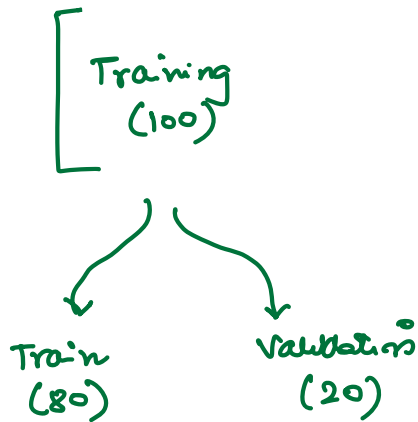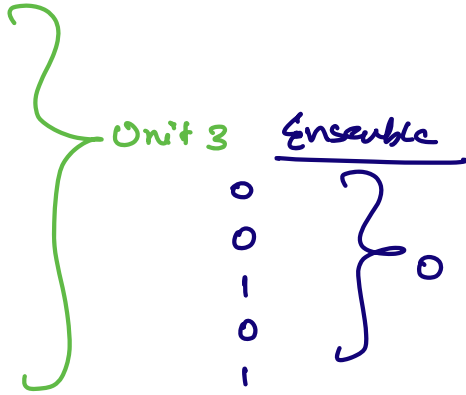
Validation (20)

Log. Regression
NB
SVM

ACC: 75%.
ACC: 80%.
Acc: 85%.

$\frac{k}{5}$ fold cross validation

100

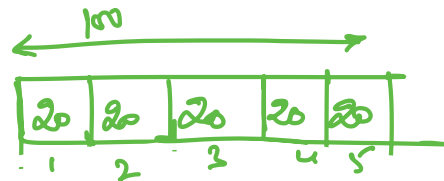| 20	20	20	20	20
1	2	3	4	5

LR
5: 1st train: 1234 Train   5th valid
   2nd train: 1235 Train   4th Valid -7
   3rd train: 1245 Train?   3rd Valid -50
   4th train: 1345 Train?   2nd Valid -
   - 5th train: 2345 Train   1st valid -30

80

Avg.

error — Training data

overfitting

epochs
↓
hyperparameter

VS

Radom w

$\left[\begin{array}{l} \text{error (MSE) loss} \\ w = \omega - \eta \dfrac{\partial L}{\partial \omega} \end{array}\right]$ → error VS

Parameter → w_1, w_2

Hyperparameter → η

epochs, knn (K)   SVM (C)
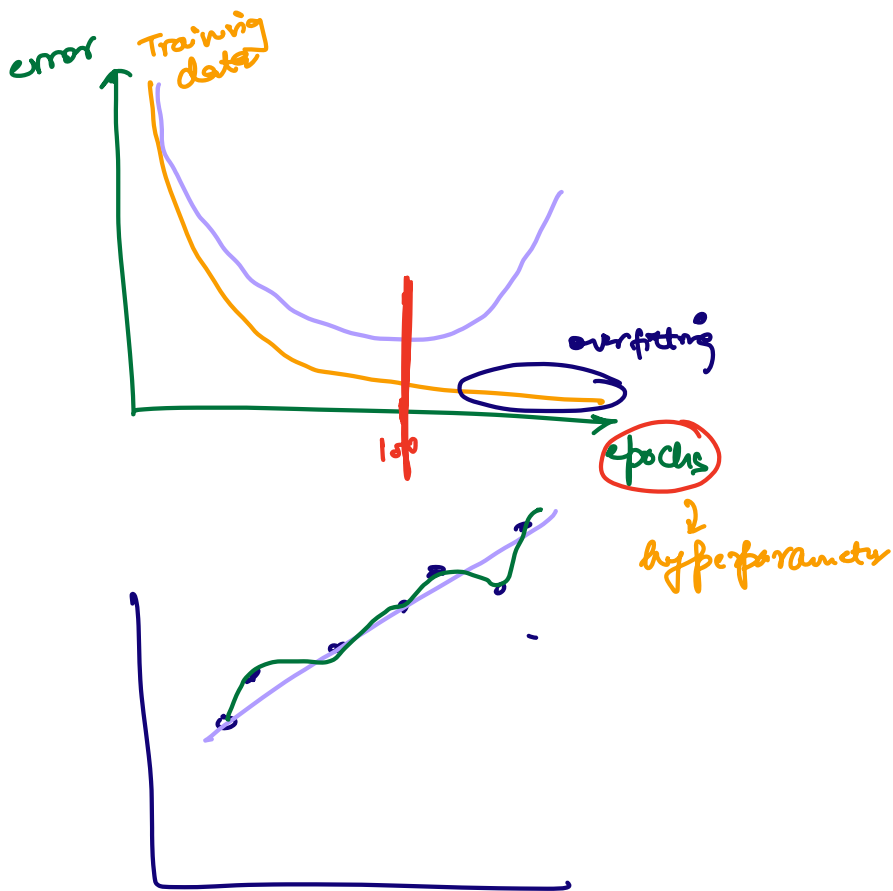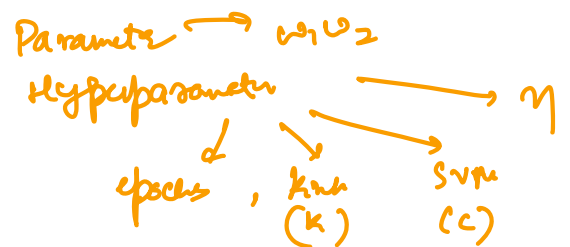
GRID SEARCH

$\eta: 0.1, 0.01, 0.05, 0.001$

best result pick
↓
validate 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	$\dfrac{9}{10} = 90\%$
0	1	0	
0	0	0	

0

1

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

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

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

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

80% → correct

→ 20% incorrect

Actual	Predicted	
0	1	} Accuracy
1	0	} fades

Confusion Matrix

Prediction

Actual	0	1
0	True Negative	False Positive
1	False Negative	True Positive

0: Not Placed
1: Placed

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

Predicted

	0	1	2
0	7	0	5
1	2	4	6
2	9	1	13

Actual

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

Imbalanced Data

→ Not terrorist

$$10,000 \nearrow \searrow$$

9999 normal   1 terrorist
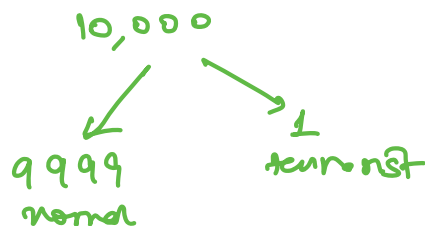
Predicted

	not terrorist 0	terrorist 1
Actual not terrorist 0	9999	0
terrorist 1	1	0

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

FP are crucial

Eg:

Predicted

		1 Spam	0 Not spam
Actual	1 Spam	100 TP	170 FN
	0 Not Spam	30 FP	700 TN

1st

Predicted

		1 Spam	0 Not spam
	1 Spam	100	190
	0 Not Spam	10	700

2nd

1st: valid mail → spam ~ 30
2nd: 10

1st: Spam → Inbox: 170
2nd: : 190

$$\text{Precision} = \frac{TP}{TP + FP} \quad \Bigg\} \quad \text{what proportion of predicted is truely correct.}$$

Precision: $\dfrac{100}{130}$

Precision: $\dfrac{100}{100+10} = \dfrac{100}{110}$

higher precise.

FP error: Type I error