

20/2/24 & 21/2/24

19

MAY'23

FRIDAY

20th Week • 139-226

44. Confusion matrix

SUN	MON	TUE	WED	THU	FRI	SAT	SUN	MON	TUE	WED	THU	FRI	SAT	SUN
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
14	15	16	17	18	19	20	21	22	23	24	25	26	27	28

Better model for classification.

it can be used for both binary & multiclass classification.

Eg) For 2 class (yes & no)

11.00
12.00
13.00
14.00
15.00
16.00
17.00
18.00

Actual \ Predicted	Yes (1)	No (0)
Yes (1)	4	1
No (0)	1	0

		Predicted (\hat{y})	
		yes (1)	no (0)
Actual	1	TP	FN
	0	FP	TN

		Predicted (P)		
		1	0	
Actual (A)	1	3 values	1 val	4
	0	2 values	1 val	3
Total value		5	2	7

It has some terminologies :-

① Accuracy

② Error

③ Precision

④ Recall

⑤ F-1 Score

⑥ AUC / ROC curve

Sensitivity

Specificity

True positive rate, False PR,

T Negative R, FNR

$$\textcircled{1} \quad \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{4}{7} = 57.1\%$$

$$\textcircled{2} \quad \text{Error} = 1 - \text{Accuracy}$$

$$= 1 - 57.1\%$$

$$= 1 - 0.57$$

$$= 43\%$$

$$\textcircled{3} \quad \text{Precision} = \frac{TP}{TP + FP}$$

True +ve rate with respect to predicted +ve.

$$= \frac{3}{5} = 0.6$$

$$= 60\%$$

$$\textcircled{4} \quad \text{Recall} = \frac{TP}{TP + FN}$$

= True Positive Rate = Sensitivity
wrt actual +ve.

$$= \frac{3}{4} = 75\%$$

note

Even after having accuracy we find recall & precision to correct the error.

$$\textcircled{5} \quad \text{F-Score} = \text{Recall} \times \text{Precision}$$

$$\textcircled{6} \quad \text{F-beta score} = 1 + (\text{beta})^2 \times \frac{(C \times P \times R)}{(\text{beta})^2 \times P + R}$$

21

MAY'23

SUNDAY

20th Week • 141-224

SUN	MON	TUE	WED	THU	FRI	SAT	SUN	MON	TUE	WED	THU	FRI	SAT	SUN
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29

SVM (Support Vector Machine)

09.00

10.00

11.00

12.00

13.00

14.00

15.00

16.00

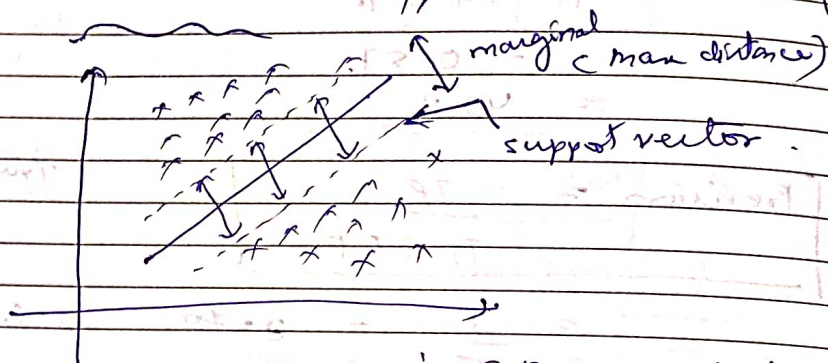
17.00

18.00

19.00

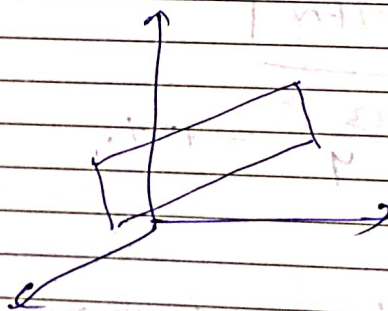
It can be used in linear reg (SVM) and in logistic reg (SVC)

It converts 2D data to 3D.

linear SVC (Support vector classifier)

in 2D we will have a line

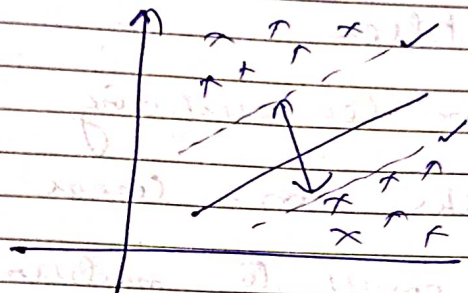
in 3D we will have a plane

Some Terminologies

- | | |
|------------------|------------------|
| ① kernel | ⑤ soft margin |
| ② kernel trick | ⑥ C Parameter |
| ③ support vector | ⑦ multiclass svm |
| ④ margin | ⑧ RBF |

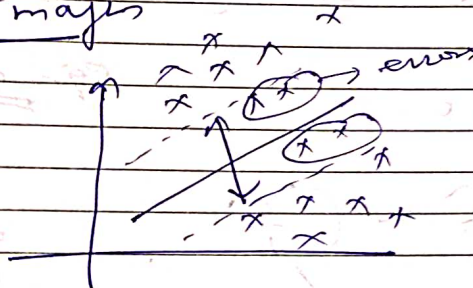
Enthusiasm contagious - it's caught not taught.

Hard margin



- ① margin dist max.
- ② parallel ✓
- ③ will not consider any error.

Soft margin



- ① parallel
- ② margin dist max
- ③ Some data points are misclassified i.e. consider error.

This can be achieved by
Cross validation

↓
 testing & training data -