Lec-26 13/03/21

Model Evaluation measures (Section 8.5, Han Kamber, 3rd Edn)

8.5.1 → metrics for evaluating classifier performance

8.5.2 → hold out method and random subsampling

8.5.3 → cross-validation

8.5.4 → bootstrap

8.5.5 → model soln. using statistical test (not reqd.)

8.5.6 → ROC curve

8.5.1 → (1) accuracy (recogn. rate)
(2) sensitivity (recall)
(3) specificity
(4) precision
(5) F1-score
(6) F_β score

Positive tuples (main class of interest) Eg: buy-comp-Yes

Negative tuples (rest all tuples) Eg: buy-comp-NO

1) True positive: (TP):

P $\xrightarrow{ML}$ P ✓

2) True Negative (TN):

N $\xrightarrow{ML}$ N ✓

3) False positive (FP)

(- +)

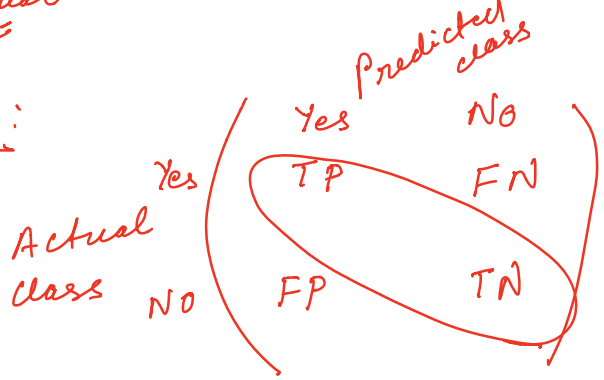
N (actual) → ML → P ✓

4) False negative (FN)

(- -)

P (actual) → ML → N

Confusion matrix :



1. accuracy (recogn. rate) = $\frac{TP + TN}{P + N}$

2. Error rate (misclassification rate): $\frac{FP + FN}{P + N}$ (or 1 - acc(M))

3. Sensitivity (TPR) (recall) $\frac{TP}{P}$

4. Specificity (TNR) : $\frac{TN}{N}$

5. precision = $\frac{TP}{TP + FP}$

6. F, F1 score, harmonic mean of precision, recall

= $\left(\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \right)$

7. F_β where β is a non-negative real no.

$$F_\beta = \frac{(1+\beta^2) \times \text{precision} \times \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}}$$

Numerical :

1) (buy. comp)

actual class	Yes	No	
Yes	6954 (TP)	46 (FN)	7000
No	412 (FP)	2588 (TN)	3000
	7366	2634	Σ = 10,000

find the i) recogn. rate ? (iv) precision

ii) sensitivity (v) Error

(iii) specificity

i) recogn rate : 95.42 % ✓

ii) sensitivity : 99.3 ✓

iii) specificity : 86.2 % ✓

iv) precision : 94.4 % ✓

v) Error : 0.045 (4.58) %

Resubstitution Error :

If the training set instead of test set is used to estimate the error rate of model, this quantity is known as resubstitution error.

Tr → model

Test → test the model

Test acc. %

→ class imbalance problem :

Fraud detection

↙ fraudulent +ve count is low 100

↘ non-fraudulent (-ve) high 1000

Cancer prediction

↙ 50 (+ve class)

↘ 1000 (normal)

50 :/. Tr 25 500

50 :/. Test 25 500

sensitivity : $\left(\frac{TP}{P}\right)$

specificity : $\frac{TN}{N}$

prob. (2)

Cancer	Yes	No	
Yes	90	210	→ 300
No	140	9560	→ 9700
	230	9770	<= 10,000

find : i) sensitivity (TPR, or recall) $\left(\frac{90}{300} = 30\text{·}/.\right)$

(ii) specificity (TNR) $\left(\frac{9560}{9700} = 98.56\text{·}/.\right)$

(iii) overall accuracy → $\frac{9650}{10,000} = 96.5/.$

(iv) precision

$\frac{90}{230} = 39.13\text{·}/.$

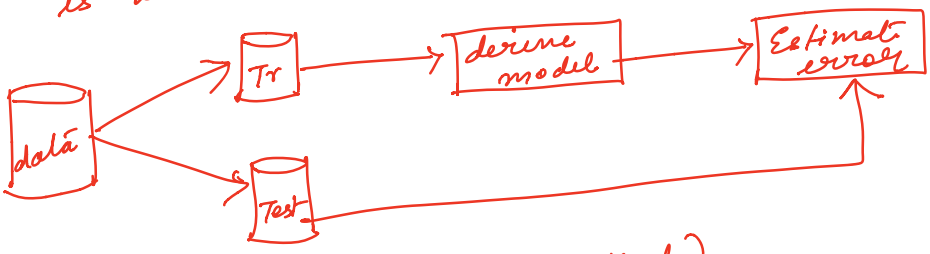
8.5.2 Hold out method and random subsampling:

Hold out: Data set

→ Test - (1/3rd) Test set → estimate the accuracy of model.

↙ Training (2/3rd dataset)

is used to derive the model



(Hold out method)

Random subsampling:

→ (repeat hold out method "K" no. of times)

→ Evaluate avg. performance of the model over K iterations.

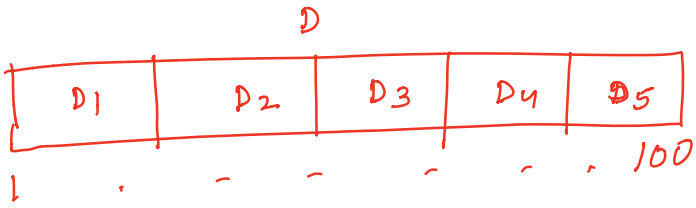
Iter. 1	60 % (M1)	40 % (A1)
Iter 2	60 % (M2)	40 % (A2)
⋮	⋮	⋮
5	(M5)	(A5)

$$\text{Over all Acc} = \frac{1}{5}\sum_{i=1}^{5} Acc_i$$

2) **cross - validation:**

K - fold CV (Assume K = 5)

5 - fold CV



Iter 1: $\{D_1\}$ → Test $\{D_2\ D_3\ D_4\ D_5\}$ Tr (M1)

2 $\{D_2\}$ Test $\{D_1\ D_3\ D_4\ D_5\}$ Tr (M2)

3 $\{D_3\}$ Test $\{D_1\ D_2\ D_4\ D_5\}$ Tr (M3)

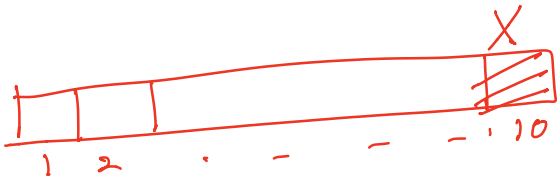
4 $\{D_4\}$ Test $\{D_1\ D_2\ D_3\ D_5\}$ Tr (M4)

5 $\{D_5\}$ Test $\{D_1\ D_2\ D_3\ D_4\}$ Tr (M5)

$$\text{overall acc} = \frac{1}{5}\sum_{i=1}^{5} \text{Acc}(M_i)$$

Leave-one-out CV:

Itm 1: 1



classifier counter

~~0 × 2~~

✓ 7 3 ×

10

Accuracy

$\left(\frac{7}{10}\right) \times 100 = 70\%$

$\left(\frac{3}{10} \times 100 = 30\%\right.$ Error$\left.\right)$