MATHEMATICAL FOUNDATIONS OF DATA ANALYSIS

UNIT IV

PERFORMANCE MEASURES

Performance measure is an idea used by the classifier to classify and analyze the given set of data points in various parameters like identification labels, overall performance of a classifier, Sensitivity, Specificity, Balanced Accuracy (BA), Prevalence (P), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Detection Rate (DR), Detection Prevalence, Observed Accuracy (OA) and Expected Accuracy (EA). The Performance Measure Table is given below.

Performance Measure Table:

	True Condition		
	Total Population	**Condition Positive**	**Condition Negative**
Predicted Condition	Predicted Condition Positive	True Positive (TP) (Success)	False Positive (FP) Type I Error (Mistake)
	Predicted Condition Negative	False Negative (FN) Type II Error (Mistake)	True Negative (TN) (Success)

Terminologies Used in Performance Measure:

TP - True Positive FP - False Positive FN – False Negative TN - True Negative

N = TP + TN + FP + FN .

TP represents correct identification of Positive Labels

TN represents correct identification of Negative Labels

FP represents incorrect identification of Positive Labels

FN represents incorrect identification of Negative Labels

Here we are listing some performance measures of a classifier.

Definition : Accuracy (A)

Accuracy (A) gives **overall effectiveness of a classifier** and is calculated by $\boldsymbol{A = \frac{TP+TN}{N}}$.
Maximum value of the Accuracy is 1. It happens when the classifier exactly classifies two groups FP = 0 and FN = 0.

Remarks:

1. Total Number of True Positive Labels = TP + FN
2. Total Number of True Negative Labels = TN + FP

Definitions:

1. **Sensitivity : Sensitivity** is defined as the effectiveness of a classifier to identify positive labels and is denoted by S_e and is calculated by using the formula
   $S_e = \frac{TP}{TP+FN}$. Here $0 < S_e < 1$.
2. **Specificity** : **Sensitivity** is defined as the effectiveness of a classifier to identify Negative labels and is denoted by S_p and is calculated by using the formula
   $S_p = \frac{TN}{FP+TN}$ and $0 < S_p < 1$.
3. The **ideal value** for S_e and S_p is 1.
4. **Balanced Accuracy (BA)** : Balanced Accuracy is defined as the average value of **Sensitivity** S_e and Specificity S_p **.**
   Therefore $BA = \frac{S_e + S_p}{2}$
5. **Prevalence (P): Prevalence** is the measure of "How often the YES condition actually occurs in the given sample **".** and is calculated by using the formula
   $P = \frac{TP+FN}{N}$
6. **Positive Predictive Value (PPV): PPV** is the **proportion of correct results in labels identified as Positive** and is calculated by using the formula.

   $$PPV = \frac{S_e \times P}{S_e \times P + (1 - S_p)(1-P)}$$
7. **Negative Predictive Value (NPV):**
   NPV is calculated by using the formula
   $$NPV = \frac{S_p \times (1 - P)}{S_p \times (1-P) + (1-S_e) \times P}$$
8. **Detection Rate (DR)** : Detection Rate (DR) is calculated by using the formula
   $DR = \frac{TP}{N}$
9. **Detection Prevalence (DP):** Detection Prevalence (DP) is calculated by using the formula $DP = \frac{TP + FP}{N}$
10. **Kappa Statistic:** Kappa Statistic is a metric that compares the Observed Accuracy (OA) with the Expected Accuracy (EA) and is calculated by using the formula .
    $Kappa = \frac{OA - EA}{1-EA}$ **.** Here

a= TP	b= FP	a+b
c= FN	d= FN	c+d
a+c	b+d	a+b+c+d = N

$$\boldsymbol{A = OA = \frac{a+d}{N} = \frac{TP+TN}{N}}$$

$$\mathbf{EA} = \frac{\mathbf{(a+c)(a+b) + (b+d)(c+d)}}{\mathbf{N}}$$

Problem:

Using Performance measure table find the following for given data set points.

Performance Measure table:

a=TP (True Positive)	b = FP (False Positive)
c = FN (False Negative)	d = TN (True Negative)

Performance Measure table:

10	1
0	9

(i) A – Accuracy (ii) Se – Sensitivity (iii) Sp – Specificity (iv) P – Prevalence (v) PPV – Positive Predictive Value .

Solution: We are given that a = TP = 10. b = FP = 1. c = FN = 0 and d = TN = 9 and N = a+b+c+d = 10+1+0+9 = 20.

(i) Accuracy $A = \frac{\text{TP+TN}}{\text{N}} = \frac{10+9}{20} = \frac{19}{20} = 0.95$

(ii) Sensitivity $S_e = \frac{TP}{TP+FN} = \frac{10}{10+0} = 1$

(iii) Specificity $S_p = \frac{TN}{FP+TN} = \frac{9}{1+9} = \frac{9}{10} = 0.9$

(iv) Prevalence $P = \frac{TP+FN}{N} = \frac{10+0}{20} = \frac{10}{20} = 0.5$

(v) Positive Predictive Value

$$PPV = \frac{S_e \times P}{S_e \times P + (1-S_p)(1-P)}$$

$$= \frac{1 \times 0.5}{1 \times 0.5 + (1-0.9)(1-0.5)} = \frac{0.5}{0.5+0.1\times0.5} = \frac{0.5}{0.55} = 0.9091$$

(vi) Negative Predictive Value

$$NPV = \frac{S_p \times (1-P)}{S_p \times (1-P) + (1-S_e) \times P}$$

$$= \frac{0.9 \times (1-0.5)}{0.9\,(10.5) + ?(1-1)0.5} = 1$$

(vi) Detection Rate $DR = \frac{TP}{N} = \frac{10}{20} = 0.5$

(vii) Detection Prevalence $DP = \frac{TP+FP}{N} = \frac{(10+1)}{20} = \frac{11}{20} = 0.55$

(viii) Balanced Accuracy $BA = \frac{S_e+S_p}{2} = \frac{1+9}{2} = 0.5$

(ix) Observed Accuracy $\text{OA} = \frac{\text{a+d}}{\text{N}} = \frac{10+9}{20} = \frac{19}{20} = 0.95$

(x) Expected Accuracy $\text{EA} = \frac{(\text{a+c})(\text{a+b}) + (\text{b+d})(\text{c+d})}{\text{N}}$

$$= \frac{(10+0)(10+1)+(1+9)(0+9)}{20} = \frac{10\times11+10\times9}{20} = \frac{200}{20} = 10$$

(xi) Kappa $Kappa = \frac{OA - EA}{1-EA} = \frac{(0.95-10)}{(1-10)} = -\frac{9.05}{-9} = 1.005$.