Classification – Performance metrics

- Accuracy
- Sensitivity
- Specificity
- Precision
- Recall
- F1 measure
- True positive rate (TPR)
- False negative rate (FNR)
- ROC
- AUC

Confusion Matrix

- Square matrix that contains all the possible classes in both the horizontal and vertical directions and list the classes along the top of a table as the predicted outputs, and then down the left-hand side as the targets.
- Example :

	Outputs		
	C_1	C_2	C_3
C_1	5	1	0
C_2	1	4	1
C_3	2	0	4

- Accuracy?

Confusion matrix – Binary class

- Confusion matrix of binary class

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)

- Accuracy ???

Confusion matrix – Binary class with metrics

		Predicted Class: Positive	Predicted Class: Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

Sensitivity and Specificity

- Sensitivity (also known as the true positive rate) is the ratio of the number of correct positive examples to the number classified as positive
- Specificity is the same ratio for negative examples.

$$\text{Sensitivity} = \frac{\#TP}{\#TP + \#FN}$$

$$\text{Specificity} = \frac{\#TN}{\#TN + \#FP}$$

Precision and Recall

- Precision is the ratio of correct positive examples to the number of actual positive examples
- **Recall** is the ratio of the number of correct positive examples out of those that were classified as positive, which is the same as **sensitivity**

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

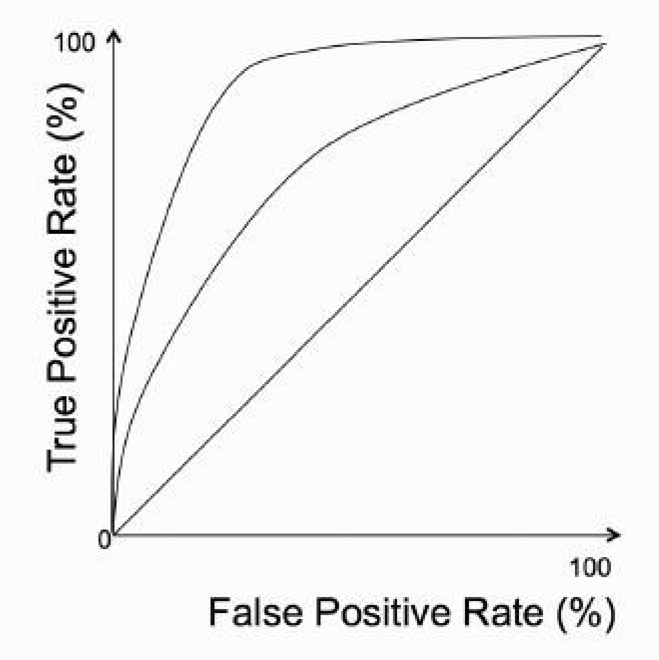
$$\text{Recall} = \frac{\#TP}{\#TP + \#FN}$$

F1 measure

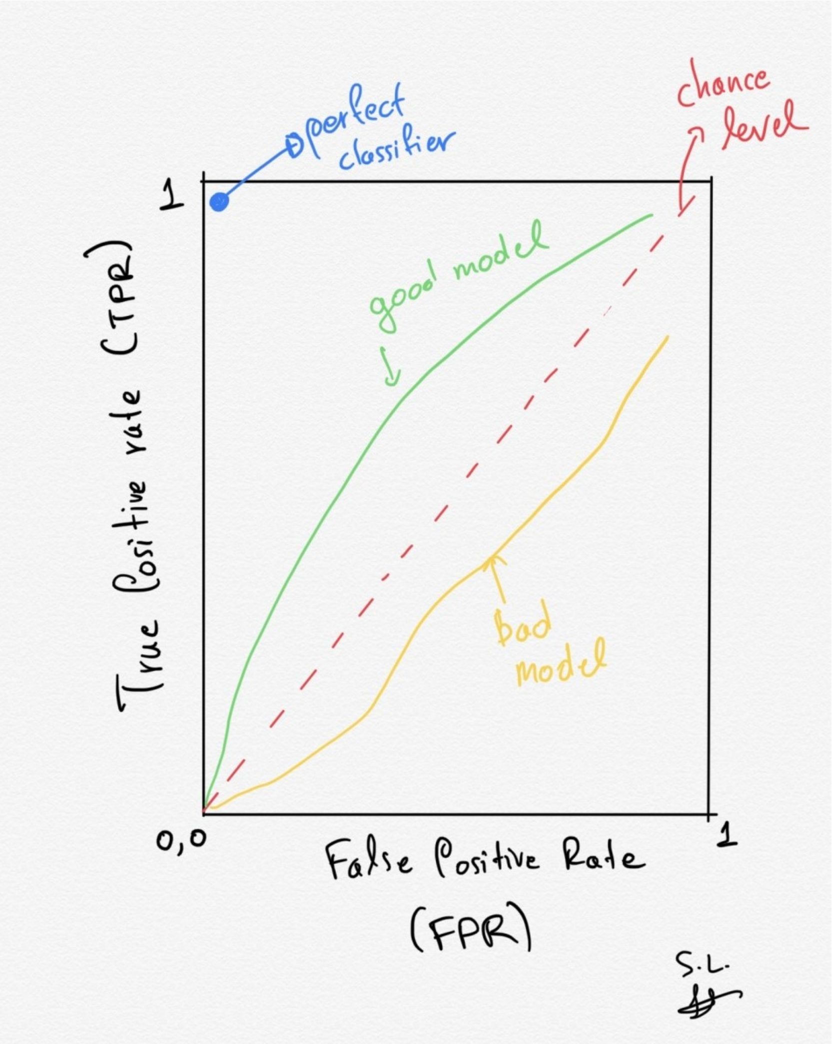
- Precision and recall are to some extent inversely related, in that if the number of false positives increases, then the number of false negatives often decreases, and vice versa.
- They can be combined to give a single measure, the *F1 measure, which can* be written in terms of precision and recall as:

$$F_1 = 2\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Receiver Operator Characteristic (ROC) Curve



Receiver Operator Characteristic (ROC) Curve



ROC - AUC

- TPR - Sensitivity
- FPR - (1-speficity)
- With imbalanced datasets, the Area Under the Curve (AUC) score is calculated from ROC and is a very useful metric in imbalanced datasets.
- The AUC (area under the curve) indicates if the curve is above or below the diagonal (chance level). AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0 and one whose predictions are 100% correct has an AUC of 1.0.

Performance metrics – overall

		True condition				
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$	F_1 score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$		

https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Accuracy Metrics – overall - Example

<table>
<tr><td colspan="2"></td><td colspan="2">Patients with bowel cancer
(as confirmed on endoscopy)</td><td colspan="3"></td></tr>
<tr><td colspan="2"></td><td>Condition positive</td><td>Condition negative</td><td>Prevalence
= (TP + FN) / Total_Population
= (20 + 10) / 2030
≈ 1.48%</td><td colspan="2">Accuracy (ACC) =
(TP + TN) / Total_Population
= (20 + 1820) / 2030
≈ 90.64%</td></tr>
<tr><td rowspan="2">Fecal occult blood screen test outcome</td><td>Test outcome positive</td><td>True positive
(TP) = 20
(2030 × 1.48% × 67%)</td><td>False positive
(FP) = 180
(2030 × (100 − 1.48%) × (100 − 91%))</td><td>Positive predictive value (PPV), Precision
= TP / (TP + FP)
= 20 / (20 + 180)
= 10%</td><td colspan="2">False discovery rate (FDR)
= FP / (TP + FP)
= 180 / (20 + 180)
= 90.0%</td></tr>
<tr><td>Test outcome negative</td><td>False negative
(FN) = 10
(2030 × 1.48% × (100 − 67%))</td><td>True negative
(TN) = 1820
(2030 × (100 − 1.48%) × 91%)</td><td>False omission rate (FOR)
= FN / (FN + TN)
= 10 / (10 + 1820)
≈ 0.55%</td><td colspan="2">Negative predictive value (NPV)
= TN / (FN + TN)
= 1820 / (10 + 1820)
≈ 99.45%</td></tr>
<tr><td colspan="2" rowspan="2"></td><td>TPR, Recall, Sensitivity
= TP / (TP + FN)
= 20 / (20 + 10)
≈ 66.7%</td><td>False positive rate (FPR), Fall-out, probability of false alarm
= FP / (FP + TN)
= 180 / (180 + 1820)
= 9.0%</td><td>Positive likelihood ratio (LR+)
$= \frac{TPR}{FPR}$
= (20 / 30) / (180 / 2000)
≈ 7.41</td><td rowspan="2">Diagnostic odds ratio (DOR)
$= \frac{LR+}{LR-}$
≈ 20.2</td><td rowspan="2">F_1 score =
$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
≈ 0.174</td></tr>
<tr><td>False negative rate (FNR), Miss rate
= FN / (TP + FN)
= 10 / (20 + 10)
≈ 33.3%</td><td>Specificity, Selectivity, True negative rate (TNR)
= TN / (FP + TN)
= 1820 / (180 + 1820)
= 91%</td><td>Negative likelihood ratio (LR−)
$= \frac{FNR}{TNR}$
= (10 / 30) / (1820 / 2000)
≈ 0.366</td></tr>
</table>

https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Check your understanding…

	Predicted class POSITIVE (spam)	Predicted class NEGATIVE (normal)
Actual class POSITIVE (spam)	TRUE POSITIVE (TP) 320	FALSE NEGATIVE (FN) 43
Actual class NEGATIVE (normal)	FALSE POSITIVE (FP) 20	TRUE NEGATIVE (TN) 538

Multiclass Classification – metrics

	Predicted class POSITIVE (spam)	Predicted class NEGATIVE (ad)	Predicted class NEGATIVE (normal)
Actual class POSITIVE (spam)	TRUE POSITIVES 27	286	40
Actual class NEGATIVE (ad)	1	37	9
Actual class NEGATIVE (normal)	5	16	500

FALSE NEGATIVES

FALSE POSITIVES

TRUE NEGATIVES

Multiclass Classification – metrics

- **True Positives**, i.e. where the actual and predicted class is spam
- **False Negatives**, where the actual class is spam, and the predicted class is normal or ad
- **False Positives**, where the actual class is normal or ad, and the predicted class is spam
- **True Negatives**, where the actual class is ad or normal, and the predicted class is ad or normal. An incorrect prediction inside the negative class is still considered as a true negative

Source: https://towardsdatascience.com/confusion-matrix-and-class-statistics-68b79f4f510b

Check your understanding…

Find the value of following performance metrics

- What is contigency table?
- What is error matrix?
- Accuracy
- Sensitivity - True positive rate (TPR) - Recall
- Specificity
- Precision
- F1 measure
- False negative rate (FNR) – Miss rate