ED5340 - Data Science: Theory and Practise

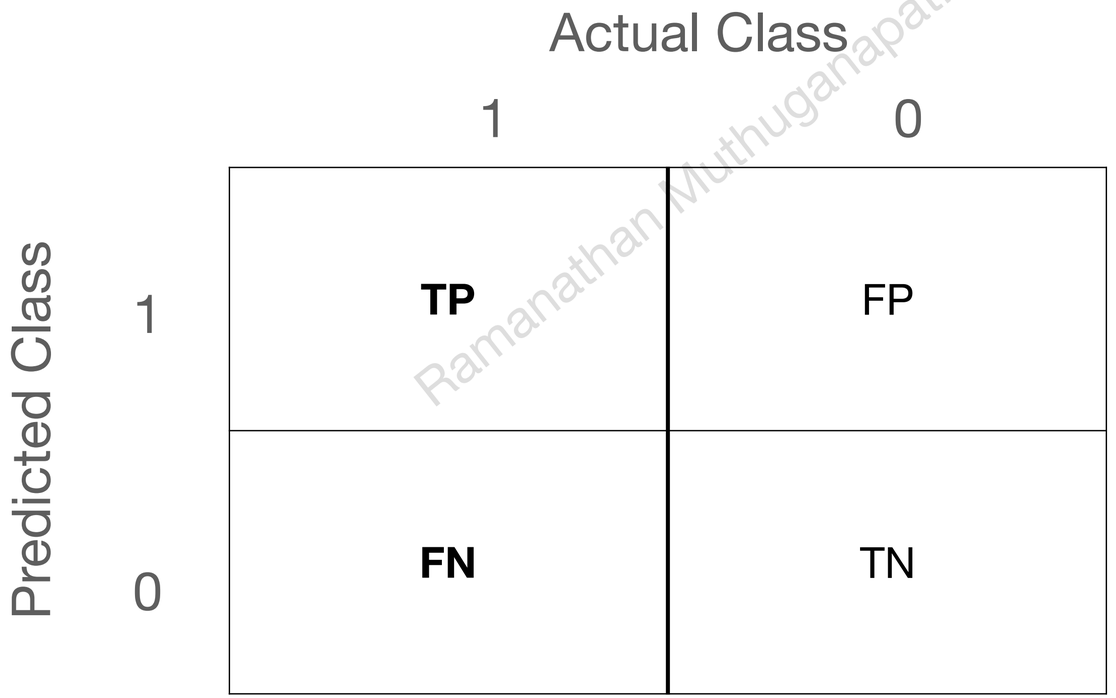
L24 - Evaluation Metrics

Ramanathan Muthuganapathy (https://ed.iitm.ac.in/~raman)
Course web page: https://ed.iitm.ac.in/~raman/datascience.html
Moodle page: Available at https://courses.iitm.ac.in/

Classification

- Confusion Matrix
- Precision
- Recall
- F_1-Score
- True positive rate
- False positive rate
- Accuracy
- AUC

Confusion Matrix



TP - True Positive

FP - False Positive

FN - False Negative

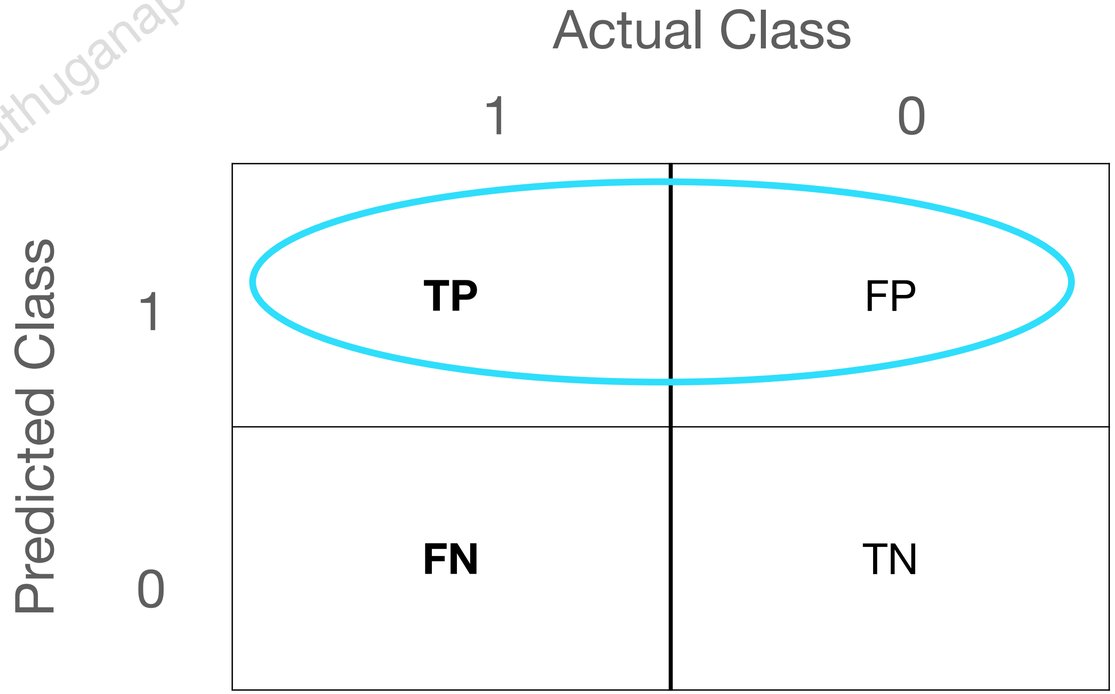
TN - True Negative

Details

- TP (True Positive) - Actual and Prediction are both positive.
- FP (False Positive) - Actual is false but the prediction is true (Prediction cancer when there is no such case).
- FN (False Negative) - Actual is true but the prediction is false (Prediction no cancer when there is one).
- TN (True Negative) - Actual and Prediction are both negative.

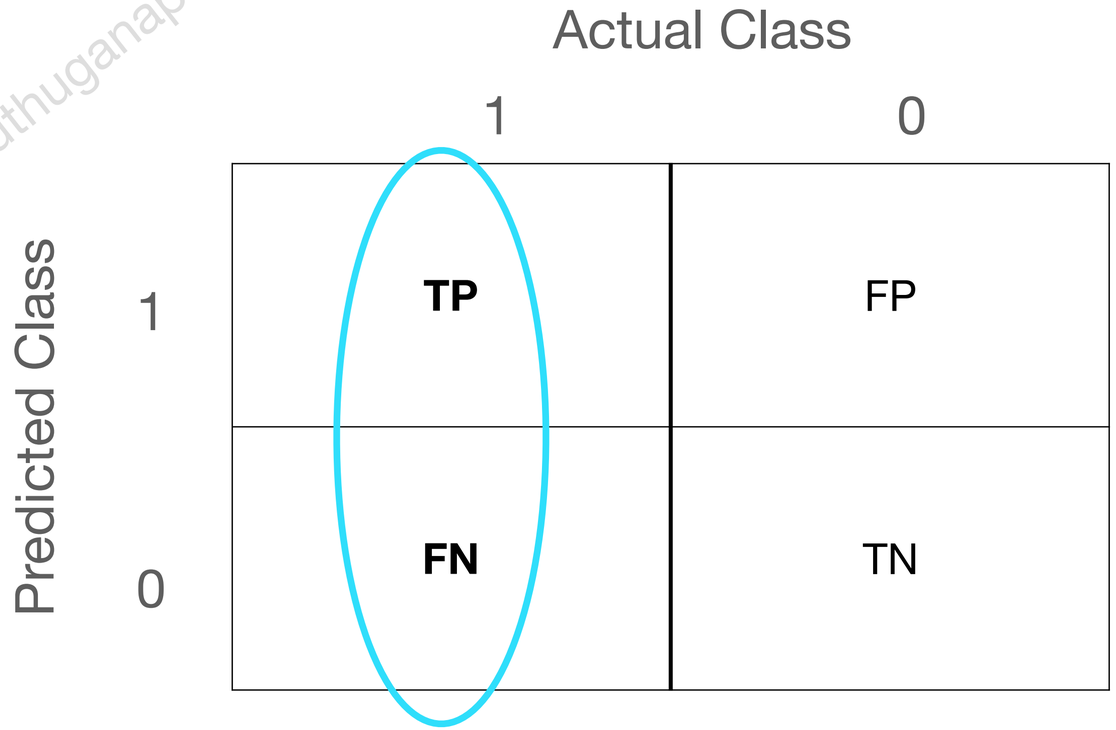
Precision and recall

- Precision - Of all the positive predicted cases, what is the fraction that is actually positive?
- $P = \frac{TP}{TP + FP}$



Precision and recall

- Recall - Of all the actual positive cases, what is the fraction that has been correctly predicted?
- $R = \frac{TP}{TP + FN}$



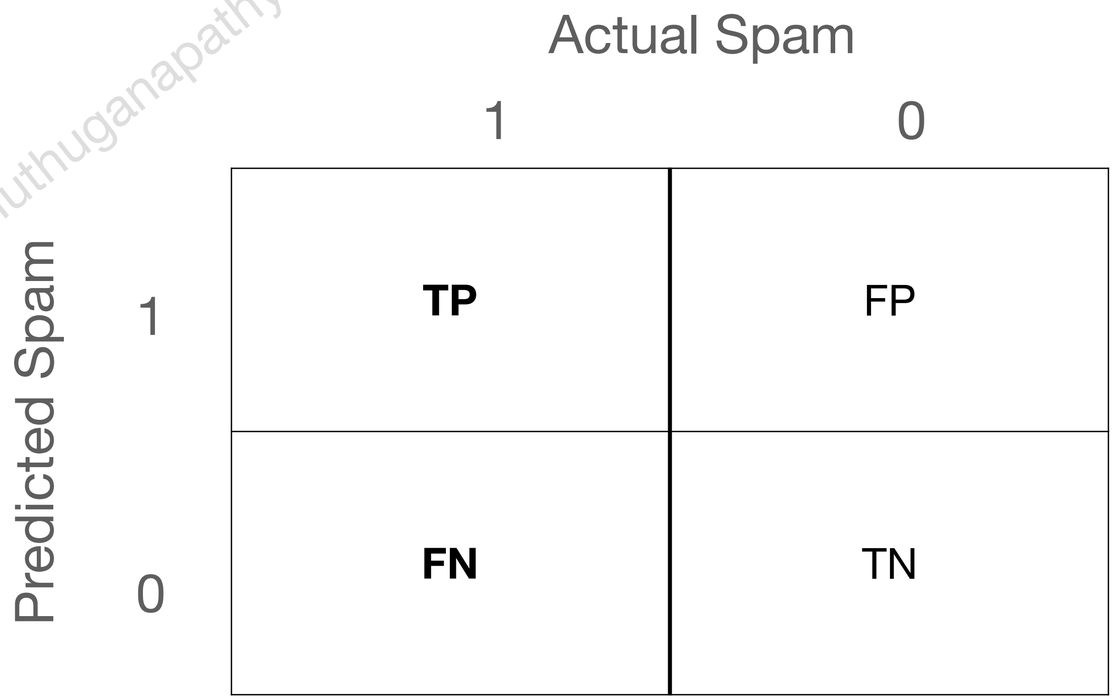
Example

- Dataset - 50 cases, 40 true and 10 false
- P = $\frac{TP}{TP + FP}$
- R = $\frac{TP}{TP + FN}$

Predicted Class \ Actual Class	1	0
1	**30**	FP
0	**FN**	3

Precision or Recall

- E.g. - Email spam filter
  - High precision or high recall
  - FP - Genuine email getting classified as spam
  - FN - Spam coming to your inbox



F_1 - Score (Harmonic mean)

- Dataset - 50 cases, 40 true and 10 false
- $F_1 = 2 * \frac{P * R}{P + R}$

Predicted Class \ Actual Class	1	0
1	**30**	FP
0	**FN**	3

Accuracy

- 50 cases, 40 true and 10 false
- Acc = $$\frac{TP + TN}{TP + FP + FN + TN}$$

Predicted Class \ Actual Class	1	0
1	**30**	FP
0	**FN**	3

TPR and FPR

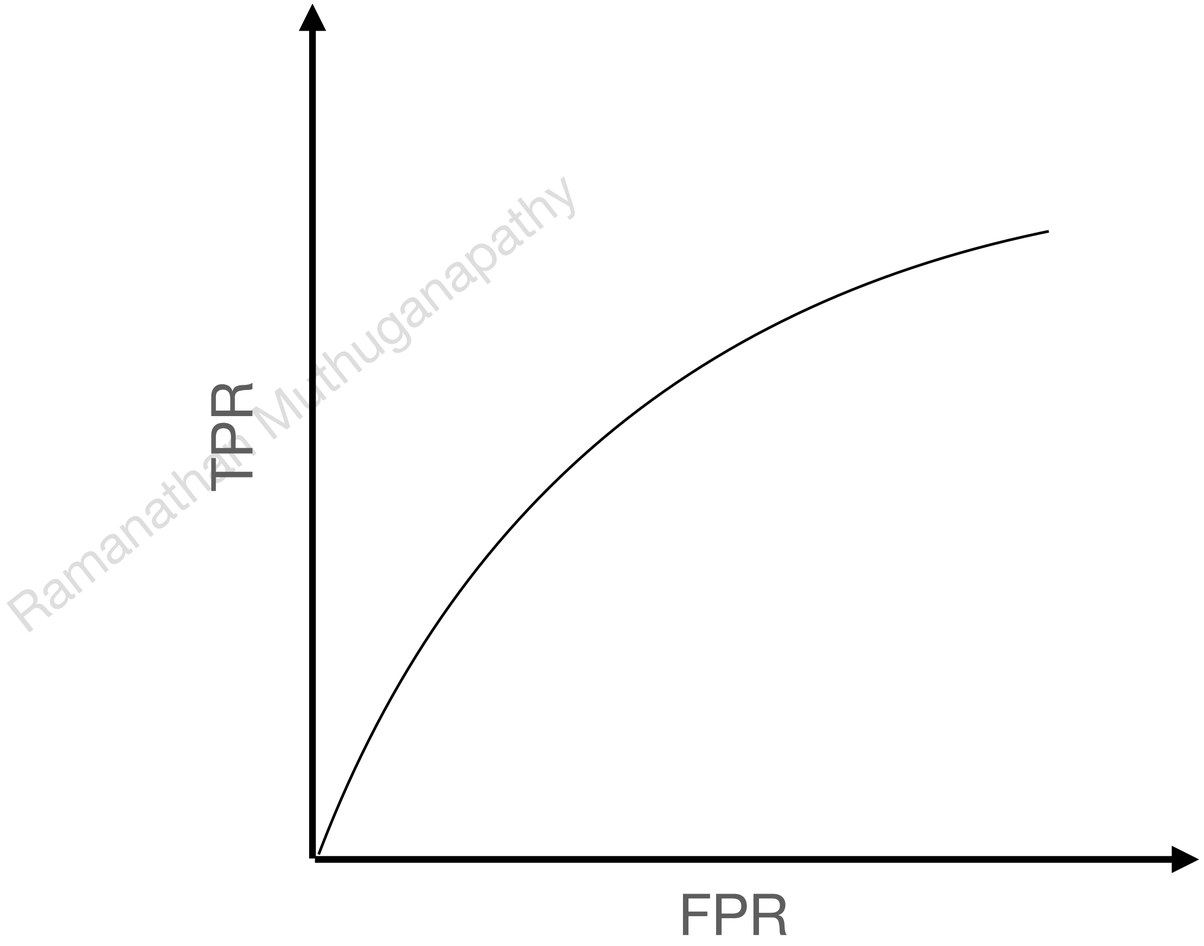
- Dataset - 50 cases, 40 true and 10 false
- TPR = $\frac{TP}{TP + FN}$
- FPR = $\frac{FP}{FP + TN}$ (negative cases being predicted incorrectly

Predicted Class \ Actual Class	1	0
1	**TP**	FP
0	**FN**	TN

Area under ROC curve (AUC)

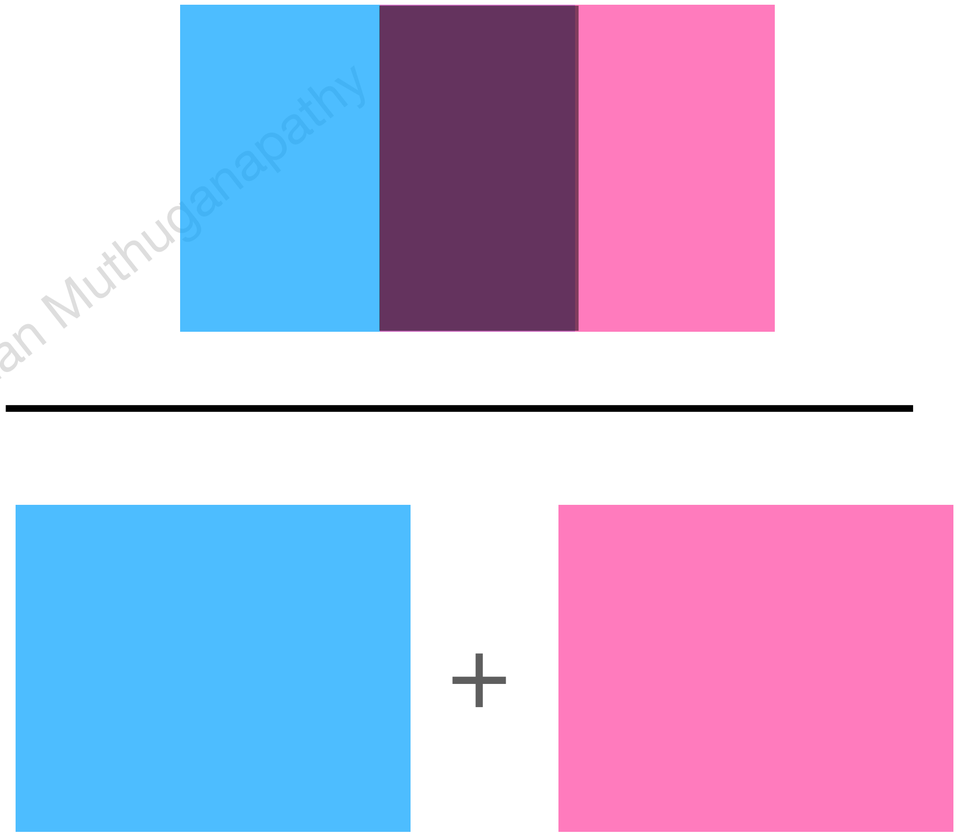
TPR and FPR

- Higher the area, the better.
- Qn: How to get this curve?



Dice score / coefficient (pixel data)

- $DC = 2 * \frac{|A \cap B|}{|A| + |B|}$
- $DC = 2 * \frac{Areaofintersection}{Sumofthetwoareas}$



IoU (Intersection over union)

- $\text{IoU} = \frac{|A \cap B|}{|A \cup B|}$
- $\text{DC} = \frac{Areaofintersection}{Areaoftheunion}$