Confusion Matrix

- used to measure performance of classification models
- n classes = n x n matrix
- Accuracy, Precision, Recall & F_1 - Score can be derived from matrix

Multi class Confusion Matrix

- TP: Diagonal of matrix
- TN, FP, FN are identified relative to a specific class of focus.

	Pred Class 1	Pred class 2	Pred class 3
Actual 1	TP_1	FN_1	FN_1
Actual 2	FP_1	TP_2	
Actual 3	FP_1		TP_3

To find TN for class 1

For class 1:

FP: $M(2,1) + M(3,1)$

FN: $M(1,2) + M(1,3)$

TN: $M(2,2) + M(2,3) + M(3,2) + M(3,3)$

↳ cover entire r1 & c1

Accuracy (Binary)

- Proportion of correctly predicted instance

$$\text{Accuracy} = \frac{TP + TN}{\text{Total Instances}}$$

- considers all predictions equally, works best when dataset balanced
- Not reliable with imbalanced dataset (eg 95% negatives, 5% positives)
  - ↳ Any inaccurate model predicting all negatives will still yield 95% accuracy

Precision (Binary)

- Measures number of instance predicted as positives are true positives.

$$\text{Precision} = \frac{TP\ (pred)}{TP(pred) + FP(pred)}$$

- focuses on quality of positive predictions
- High precision = few false positives.
- Impt when false positives have high cost

  eg: Email spam detection → FP of legitimate email causes problem

- ignores false negatives.

Recall (Sensitivity) (Binary)

- measures how many of actual positive instance are correctly predicted as positive.

$$\text{Recall} = \frac{TP\ (pred)}{TP(pred) + FN(\text{Actl True})}$$

- emphasizes on finding all positives
- Impt when missing a positive have high cost

  eg: Medical diagnosis → missing patient

- High Recall leads to lower Precision

  ↳ if ↑ FP, FN ↓

F1 - Score (Binary)

- Harmonic mean of precision & Recall

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- High F_1 score = High Precision & Recall
- Does not consider TN.

Accuracy (multi-class)

$$Acc = \frac{\text{correct prediction across all classes}}{\text{Total prediction}}$$ ← compare b/w $\hat{y}_i = y_i$

← N : No of data.

Precision, Recall, F1-score (multi-class)

- Calculated per class then aggregated
  - ↳ micro-averaging : Treat all instances as one pooled dataset
  - ↳ macro-averaging : compute metrics for each class independently then averaged

eg:

	Pred A	Pred B	Pred C
Actual A	50	2	3
Actual B	5	40	5
Actual C	2	5	45

$$\text{Precision A} = \frac{TP}{TP+FP} = \frac{50}{50+5+2} \approx 0.877 \qquad \text{Precision B} = \frac{40}{40+2+5} \approx 0.851$$

$$\text{Recall A} = \frac{TP}{TP+FN} = \frac{50}{50+2+3} = 0.909 \qquad \text{Recall B} = \frac{40}{40+5+5} \approx 0.8$$

$$F_1\text{-score} = 2\frac{(0.909)(0.877)}{0.909+0.877} \approx 0.893 \qquad F_1\text{-Score} = 2\frac{(0.8)(0.851)}{0.8+0.851} \approx 0.825$$

$$\text{Precision C} = \frac{45}{45+3+5} \approx 0.849$$

$$\text{Recall C} = \frac{45}{45+2+5} \approx 0.865$$

$$F_1\text{-score} = 2\frac{(0.849)(0.865)}{0.849+0.865} = 0.857$$

Micro-Averaging

$$TP_G = TP_A + TP_B + TP_C$$

$$\text{Precision}_G = \frac{TP_G}{TP_G + FP_G} = \frac{50+40+45}{135 + \underbrace{(7+7+8)}_{FP_G}}$$

Macro-Averaging

$$\frac{P_A + P_B + P_C}{3}$$