Model Evaluation

INSAID

Course Overview

You are here...

Term	CDF	GCD	GCDAI	PGPDSAI
Term 1	Data Analytics with Python	Data Analytics with Python	Data Analytics with Python	Data Analytics with Python
Term 2	Data Visualization Techniques	Data Visualization Techniques	Data Visualization Techniques	Data Visualization Techniques
Term 3	EDA & Data Storytelling	EDA & Data Storytelling	EDA & Data Storytelling	EDA & Data Storytelling
		Minor Project	Minor Project	Minor Project
Term 4		Machine Learning Foundation	Machine Learning Foundation	Machine Learning Foundation
Term 5		Machine Learning Intermediate	Machine Learning Intermediate	Machine Learning Intermediate
Term 6		Machine Learning Advanced (Mandatory)	Machine Learning Advanced (Mandatory)	Machine Learning Advanced (Mandatory)
		Data Visualization with Tableau (Elective - I)	Data Visualization with Tableau (Elective - I)	Data Visualization with Tableau (Elective - I)
		Data Analytics with R (Elective - II)	Data Analytics with R (Elective - II)	Data Analytics with R (Elective - II)
		Capstone Project	Capstone Project	Capstone Project
Term 7		Bonus: Industrial ML (ML – 4 & 5)	Basics of AI, TensorFlow, and Keras	Basics of AI, TensorFlow, and Keras
Term 8			Deep Learning Foundation	Deep Learning Foundation
Term 9			NPL – I/CV – I	CV – I
Term 10			NLP – II/CV – II	NLP – I
			Capstone Project	Capstone Project
Term 11				CV – II
Term 12				NLP – II
				NLP – III + CV – III
				AutoVision & AutoNLP
				Building AI product

INSAID

Term Context

- Introduction to Machine Learning
- Linear Regression
- Logistic Regression
- **Model Evaluation Techniques**  You are here…

INSAID

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- R – Squared
- Adjusted R – Squared

Classification Measures

- **Confusion Matrix**
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- Classification Report

INSAID

1. Confusion Matrix

- It **describes** the **performance summary** of prediction result of a classification model.
- Gives types of errors being made i.e. **Type I** and **Type II Errors**.

(General View)

	Predicted Negative	Predicted Positive	Total
Actual Negative	True Negative	False Positive	N
Actual Positive	False Negative	True Positive	P
Total	N'	P'	P + N

(**Example:** Cancer Classification)

	Predicted No	Predicted Yes	Total
Actual No	100	10	110
Actual Yes	5	50	55
Total	105	60	165

False Positive = Type 1 Error
False Negative = Type 2 Error

INSAID

Agenda

- Need of Model Evaluation
- Regression Measures
  - Mean Absolute Error (MAE)
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)
  - Mean Absolute Percentage Error(MAPE)
  - R – Squared
  - Adjusted R – Squared
- Classification Measures
  - Confusion Matrix (GOD FATHER)
  - **Accuracy**
  - Classification Error
  - Sensitivity / Specificity
  - Precision / Recall
  - F – Measure / Fβ – Measure
  - ROC/AUC
  - Classification Report

these all r derived out of confusion matrix

INSAID

2. Accuracy

- It gives a measure of **correct classification** of tuples when the class distribution is relatively balanced.
- In the pattern recognition literature, this is **also referred** to **as** the **Overall Recognition Rate** of the classifier.

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

	Predicted No	Predicted Yes	Total
Actual No	100	10	110
Actual Yes	5	50	55
Total	105	60	165

(Relatively Balanced Class Data)

$$\text{Accuracy} = \frac{50 + 100}{50 + 100 + 10 + 5} = \frac{150}{165} = 0.93$$

INSAID

Agenda

- Need of Model Evaluation
- Regression Measures
  - Mean Absolute Error (MAE)
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)
  - Mean Absolute Percentage Error(MAPE)
  - R – Squared
  - Adjusted R – Squared
- Classification Measures
  - Confusion Matrix
  - Accuracy
  - **Classification Error**
  - Sensitivity / Specificity
  - Precision / Recall
  - F – Measure / Fβ – Measure
  - ROC/AUC
  - Classification Report

antonyms (

95%. — 90%. — 80%.

5%. — 10%. — 20%.

INSAID

3. Classification Error

- It gives a measure of **incorrect classification** of tuples when the class distribution is relatively balanced.
- This is also referred to as the **Misclassification** rate of the classifier.

$$\text{Error Rate} = \frac{FP + FN}{TP + TN + FP + FN}$$

	Predicted No	Predicted Yes	Total
Actual No	100	10	110
Actual Yes	5	50	55
Total	105	60	165

(Relatively Balanced Class Data)

$$\text{Error Rate} = \frac{10 + 5}{50 + 100 + 10 + 5} = \frac{15}{165} = 0.09$$

INSAID

Agenda

- Need of Model Evaluation
- Regression Measures
  - Mean Absolute Error (MAE)
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)
  - Mean Absolute Percentage Error(MAPE)
  - R – Squared
  - Adjusted R – Squared
- Classification Measures
  - Confusion Matrix
  - Accuracy
  - Classification Error
  - **Sensitivity / Specificity**
  - Precision / Recall
  - F – Measure / Fβ – Measure
  - ROC/AUC
  - Classification Report

INSAID

4. Sensitivity / Specificity

- Sensitivity is referred to as **True Positive Rate** i.e., the proportion of positive tuples that are correctly identified.
- Specificity is referred to as **True Negative Rate** i.e., the proportion of negative tuples that are correctly identified.

$$\text{Sensitivity} = \frac{TP}{P}$$

$$\text{Specificity} = \frac{TN}{N}$$

	Predicted No	Predicted Yes	Total
Actual No	9560	140	9700
Actual Yes	210	90	300
Total	9770	230	10,000

(Imbalanced Class Data)

Sensitivity = 90 ÷ 300 = 0.30

Specificity = 9560 ÷ 9700 = 0.98

INSAID

Agenda

- Need of Model Evaluation
- Regression Measures
  - Mean Absolute Error (MAE)
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)
  - Mean Absolute Percentage Error(MAPE)
  - R – Squared
  - Adjusted R – Squared
- Classification Measures
  - Confusion Matrix
  - Accuracy
  - Classification Error
  - Sensitivity / Specificity
  - **Precision / Recall**
  - F – Measure / Fβ – Measure
  - ROC/AUC
  - Classification Report

INSAID

5. Precision / Recall

- **Precision** provides **measure of exactness**, i.e. what percentage of tuples labelled as positive are actually such.
- **Recall** provides **measure of completeness**, i.e. what percentage of positive tuples are labelled as such.

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

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

	Predicted No	Predicted Yes	Total
Actual No	9560	140	9700
Actual Yes	210	90	300
Total	9770	230	10,000

(Imbalanced Class Data)

Precision = 90 ÷ 230 = 0.39

Recall = 90 ÷ 300 = 0.30

INSAID

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- R – Squared
- Adjusted R – Squared

Classification Measures

- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- **F – Measure / Fβ – Measure**
- ROC/AUC
- Classification Report

INSAID

6. F – Measure / F_β – Measure

- It is also known as F_1 score which is an alternative way to **combine Precision and Recall** into single measure.
- It is the **harmonic mean** of **precision and recall** by which it gives equal **weight** to precision and recall.

50%. 50%.

$$F = 2 * \frac{precision * recall}{precision + recall}$$

	Predicted No	Predicted Yes	Total
Actual No	9560	140	9700
Actual Yes	210	90	300
Total	9770	230	10,000

(Imbalanced Class Data)

$$F = 2 * \frac{0.39 * 0.30}{0.39 + 0.30} = 0.339$$

INSAID

6. F – Measure / F_β – Measure Cont.

- It is a weighted measure of precision and recall which assigns β times as much weight to recall as to precision.
- Commonly used F_β measures are F_2 and $F_{0.5}$.

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

	Predicted No	Predicted Yes	Total
Actual No	9560	140	9700
Actual Yes	210	90	300
Total	9770	230	10,000

(Imbalanced Class Data)

$$F_{0.5} = \frac{(1 + 0.5^2) * 0.39 * 0.30}{0.5^2 * 0.39 + 0.30} = 0.36$$

INSAID

Agenda

- Need of Model Evaluation
- Regression Measures
  - Mean Absolute Error (MAE)
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)
  - Mean Absolute Percentage Error(MAPE)
  - R – Squared
  - Adjusted R – Squared
- Classification Measures
  - Confusion Matrix
  - Accuracy
  - Classification Error
  - Sensitivity / Specificity
  - Precision / Recall
  - F – Measure / Fβ – Measure
  - **ROC/AUC**
  - Classification Report

INSAID

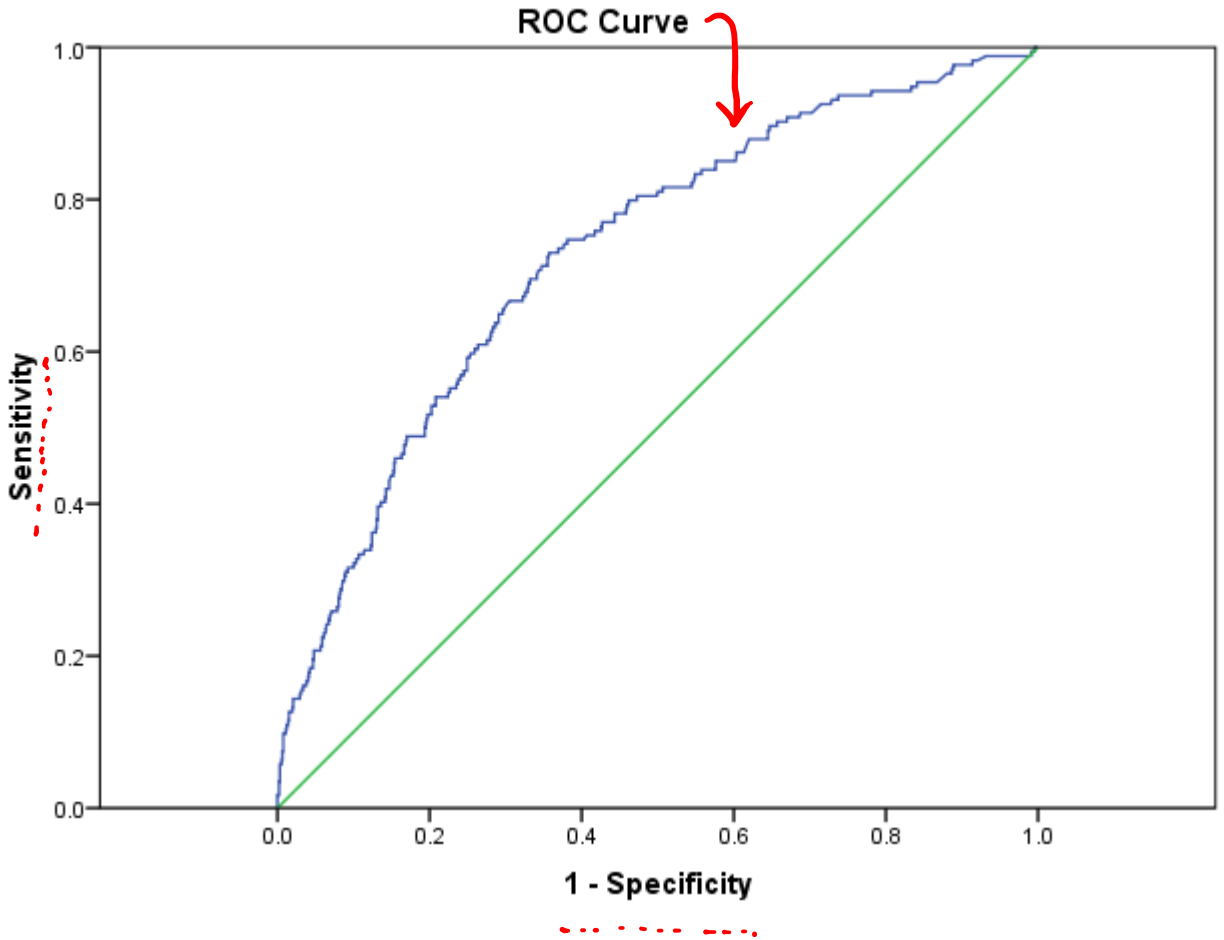
7. ROC/AUC

- Receiver Operating Characteristic (ROC)/Area Under Curve (AUC).
- It provides **visual measurement** of performance for classification problem at various **thresholds** settings.
- It is also written as **AUROC** (Area Under the Receiver Operating Characteristics).
- ROC is a **probability curve** and AUC represents **degree** or measure of **separability**.

- It tells us the **capability** of model in **distinguishing** between **classes**.
- **Higher** the **AUC**, **better** the **model** is at predicting 0s as 0s and 1s as 1s.

INSAID

Graphical Representation of AUROC



Specificity = True Negative Rate

Y → **Sensitivity** = True Positive Rate

1 – Specificity = False Positive Rate

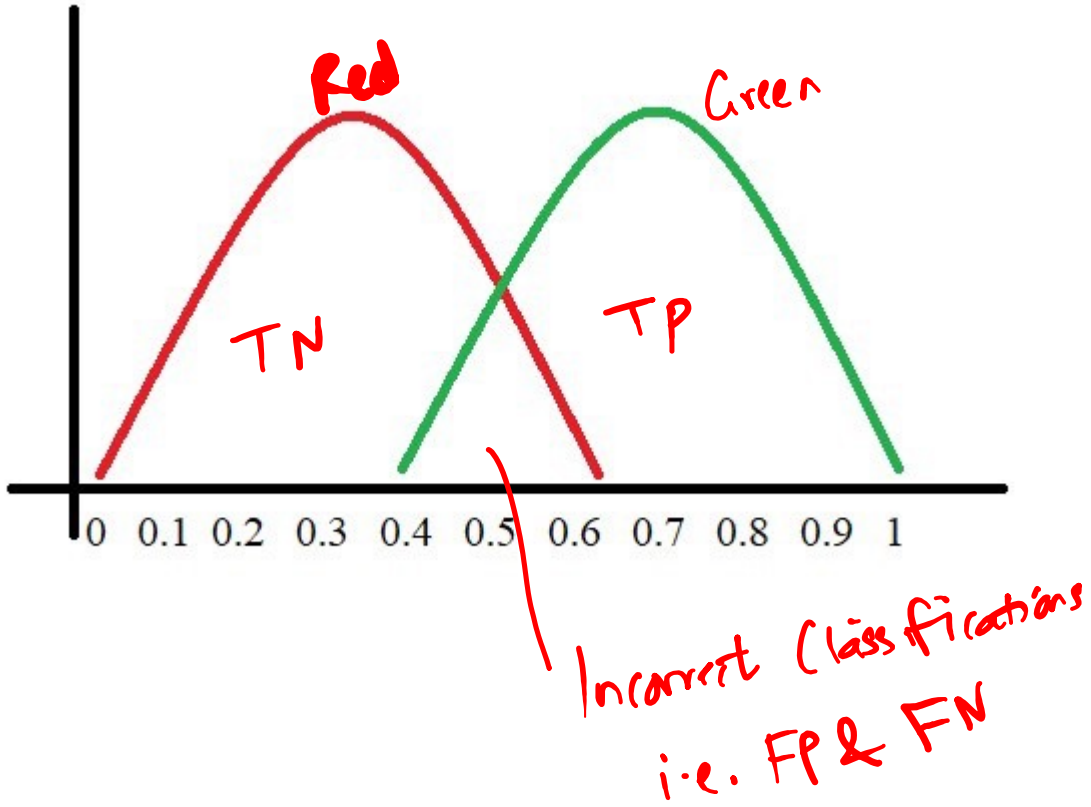
X → **1 – Specificity** = FP ÷ (TN + FP)

X vs Y → ROC Curve

INSAID

AUROC Example

- Let's say a model predicts probabilities whether a person having cancer or not.
- Using these probabilities, we plot the distribution as shown:



- **Red distribution** = Cancer No
- **Green distribution** = Cancer Yes

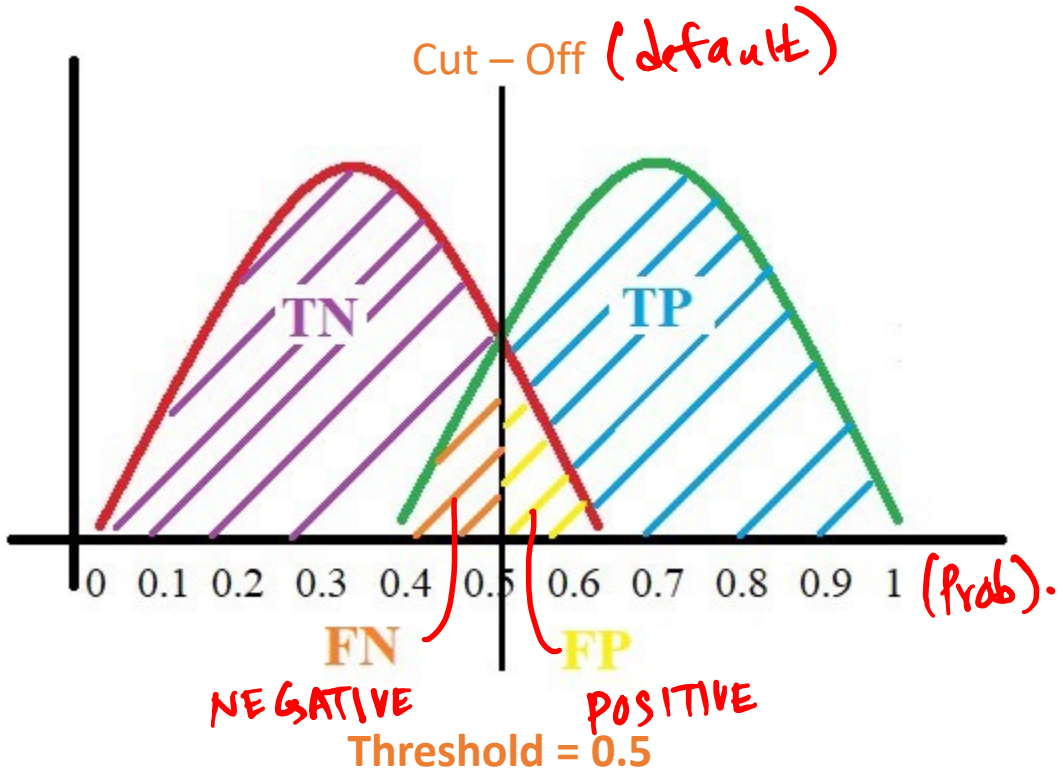
TP, TN → (max)

FP, FN → (min)

INSAID

AUROC Example Cont.

- A **cut – off** is decided to divide the patients having cancer or not. This is known as **threshold point**.

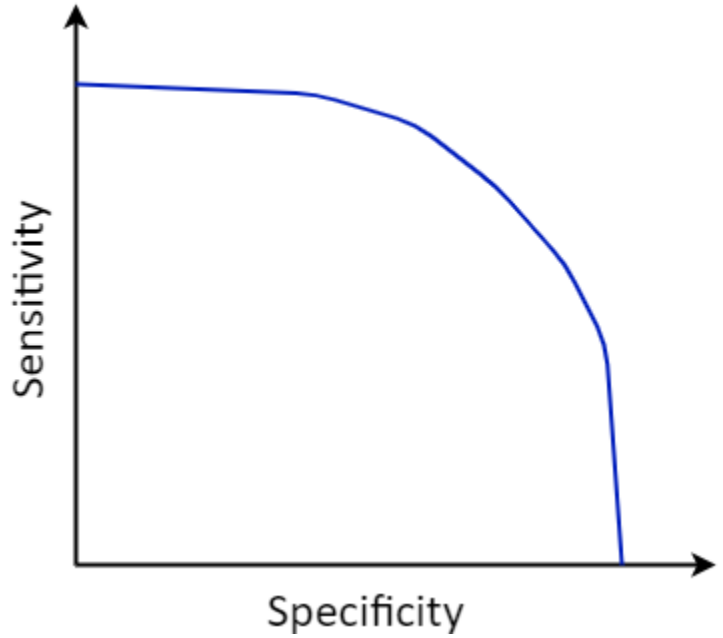


- Positive values above threshold - **True Positives**
- Negative values above threshold - **False Positives**
- Negative values below threshold - **True Negatives**
- Positive values below threshold - **False Negatives**

INSAID

Trade Off between Sensitivity vs Specificity

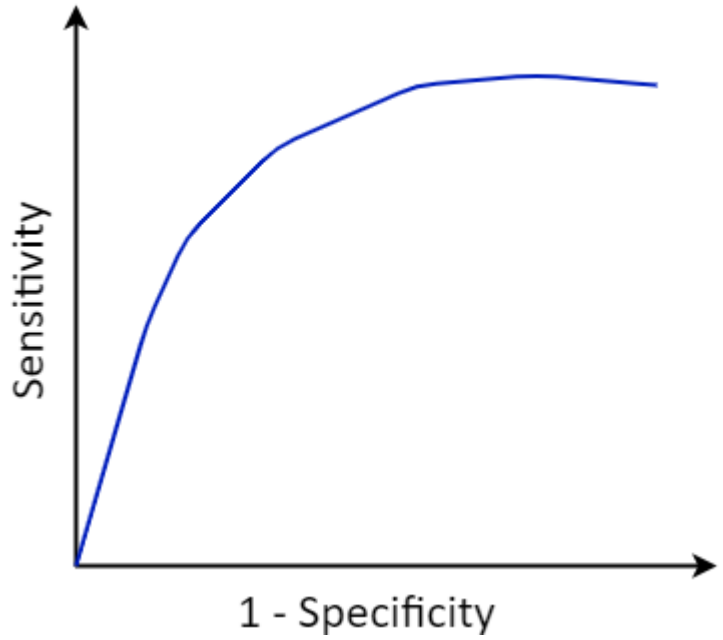
- When we **decrease** the **threshold**, we get **more positive values** i.e. **Sensitivity increases & Specificity decreases**.
- When we **increase** the **threshold**, we get **more negative values** i.e. **Sensitivity decreases & Specificity increases**.



INSAID

Trade Off between Sensitivity vs Specificity Cont.

- To plot ROC curve, instead of Specificity we use (1 — Specificity).
- Now, as we increase the threshold, we decrease the TPR as well as the FPR.
- When we decrease the threshold, we are increasing the TPR and FPR.

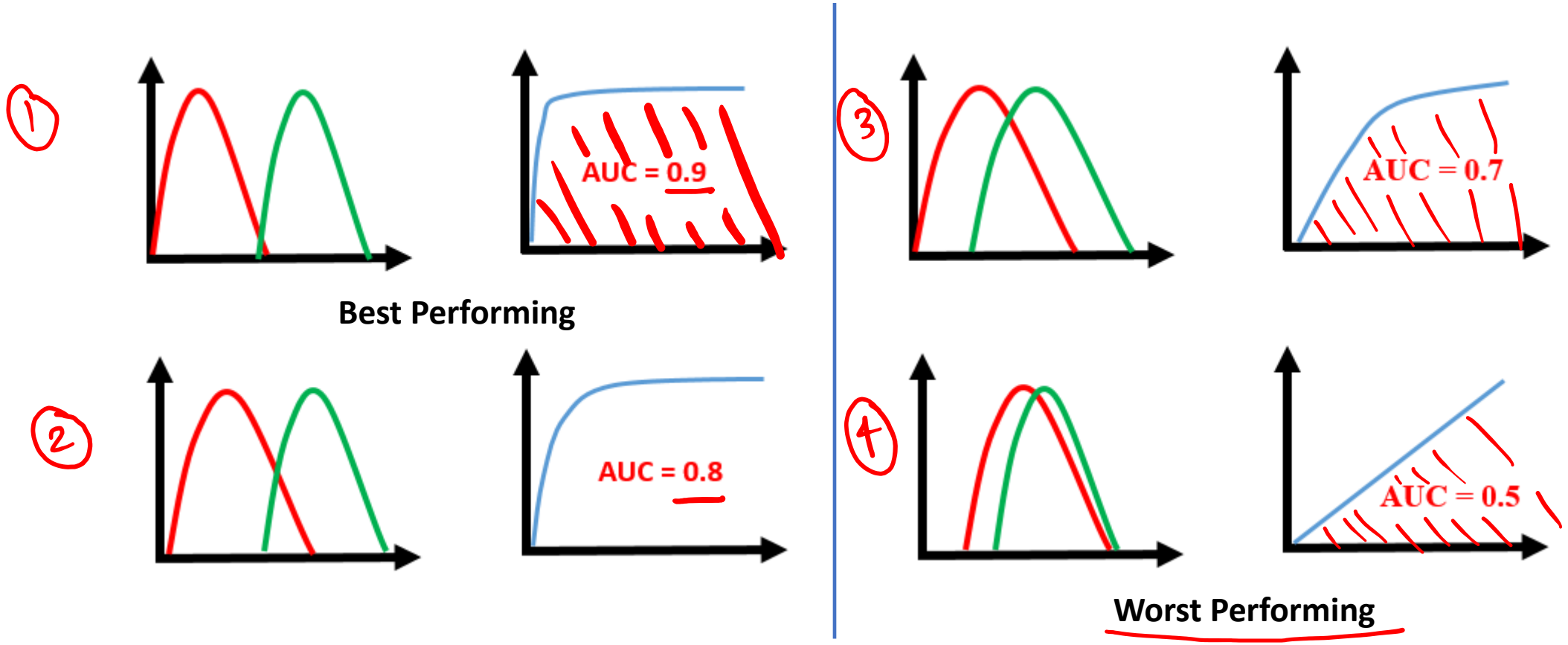


- As **Sensitivity increases**, **1 – Specificity also increases**.
- As **Sensitivity decreases**, **1 – Specificity also decreases**.

INSAID

Area Under Curve (AUC)

- The AUC is the area under the ROC curve.
- This score gives us a good idea of how well the model performances.

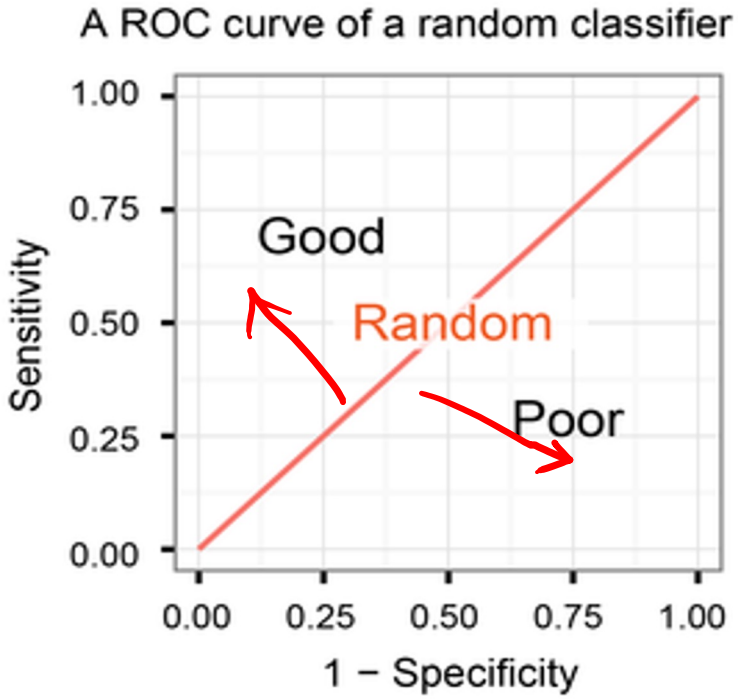


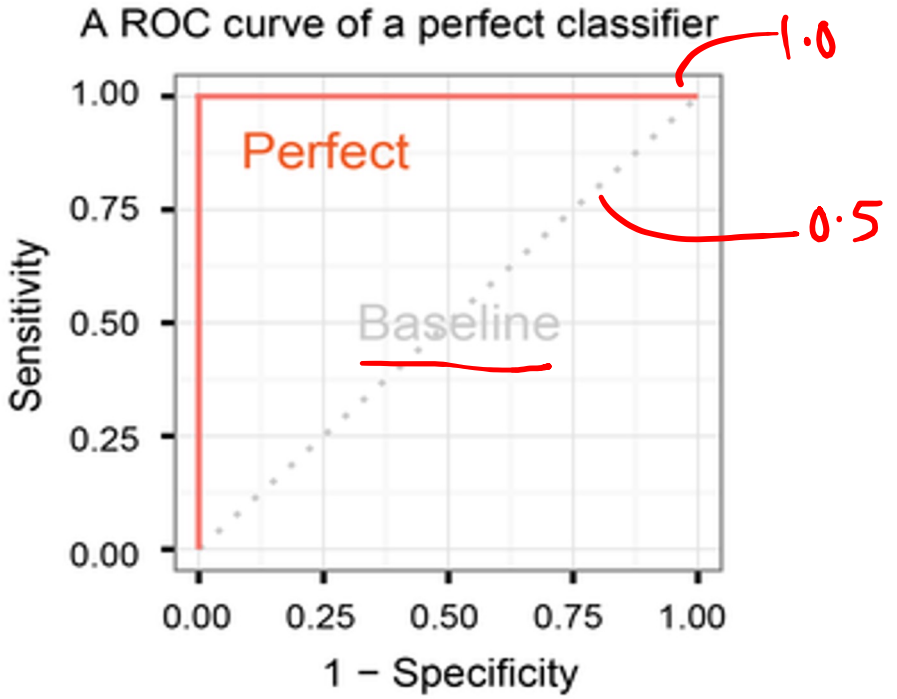
INSAID

ROC – AUC Score Comparison

- Classifiers with good performance usually lie in between the random ROC curve (baseline) and the perfect ROC curve.

Acceptable val. = min 0.75 or 0.80



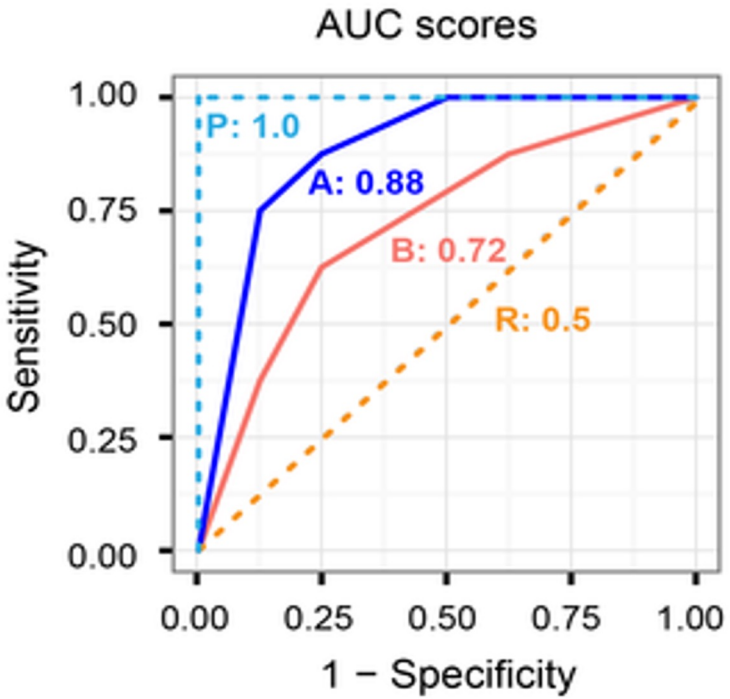


INSAID

Model Comparisons

- Model P with AUC = 1 gives perfect performance.
- Model R having AUC = 0.5 gives random predictions.
- Classifier A outperforms classifier B, as AUC score of A > B

P > A > B > R
best worst



INSAID

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- R – Squared
- Adjusted R – Squared

Classification Measures

- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- **Classification Report**

INSAID

8. Classification Report

- The classification report function **builds a text report** showing the main classification metrices.
  - **Precision**
  - **Recall**
  - **F1-score**
  - **Support**: Support is the number of occurrences of each class in Actual Target.

	precision	recall	f1-score	support
0	0.77	0.89	0.82	106
1	0.79	0.62	0.69	73
avg / total	0.78	0.78	0.77	179

Classification Report of the Cancer Survival model

Till Next Time…

INSAID