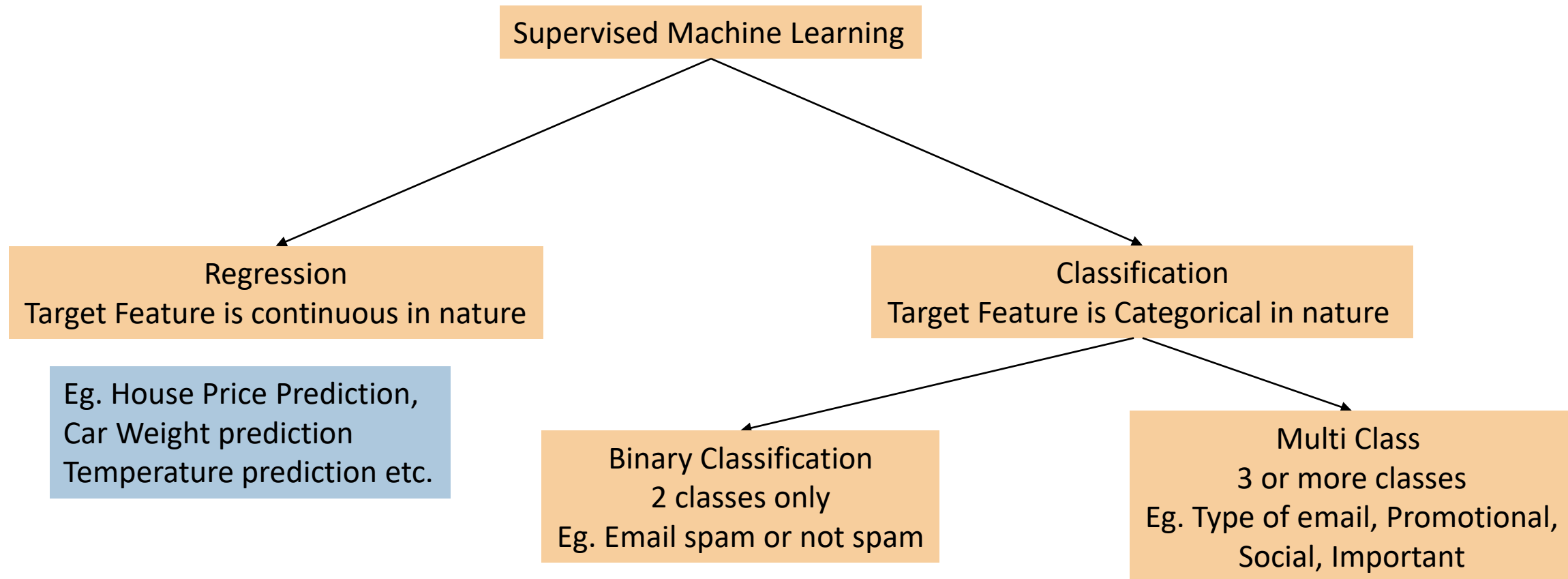


Classification

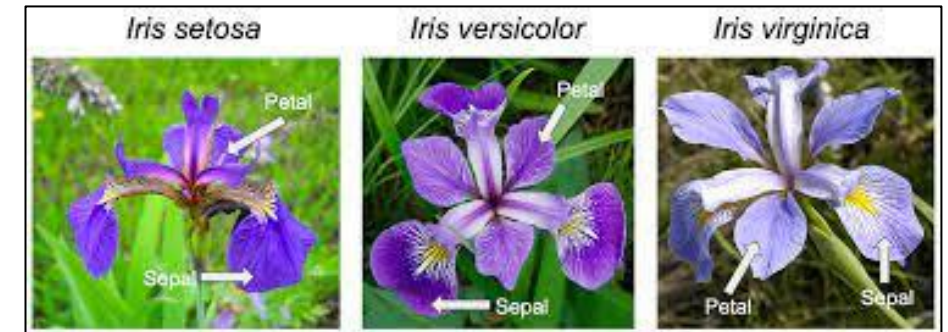
UTKARSH GAIKWAD

Types of Supervised machine learning

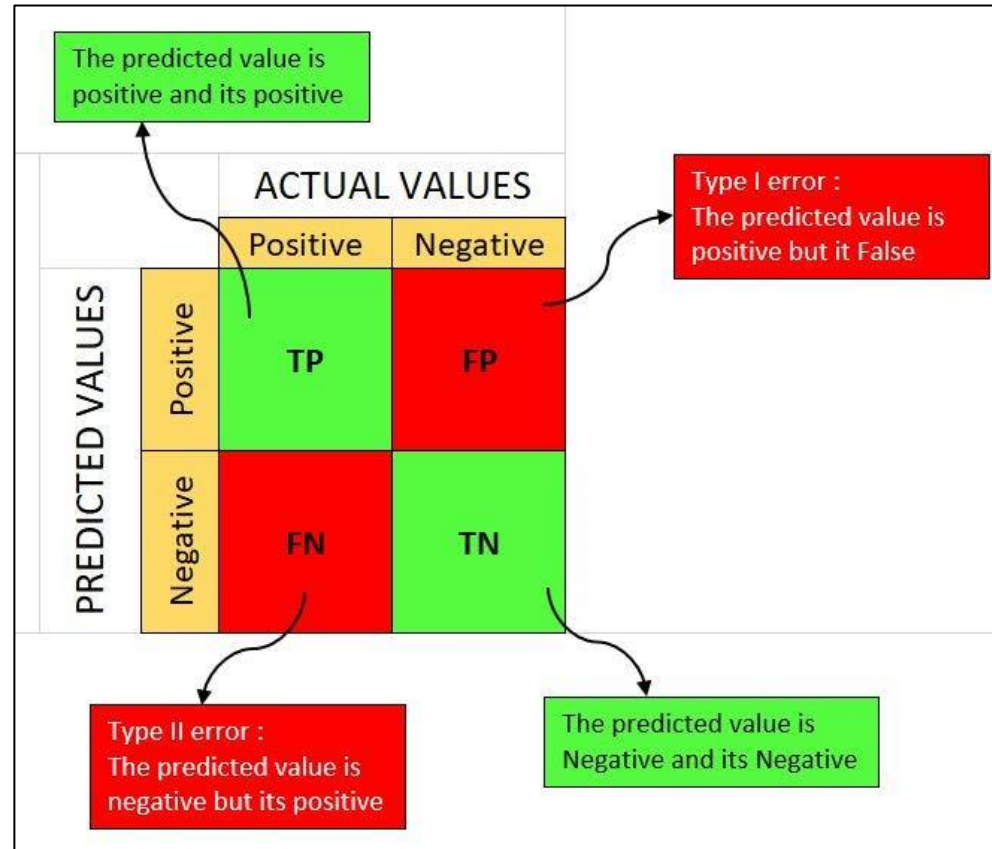


Example of classification Problem IRIS Data

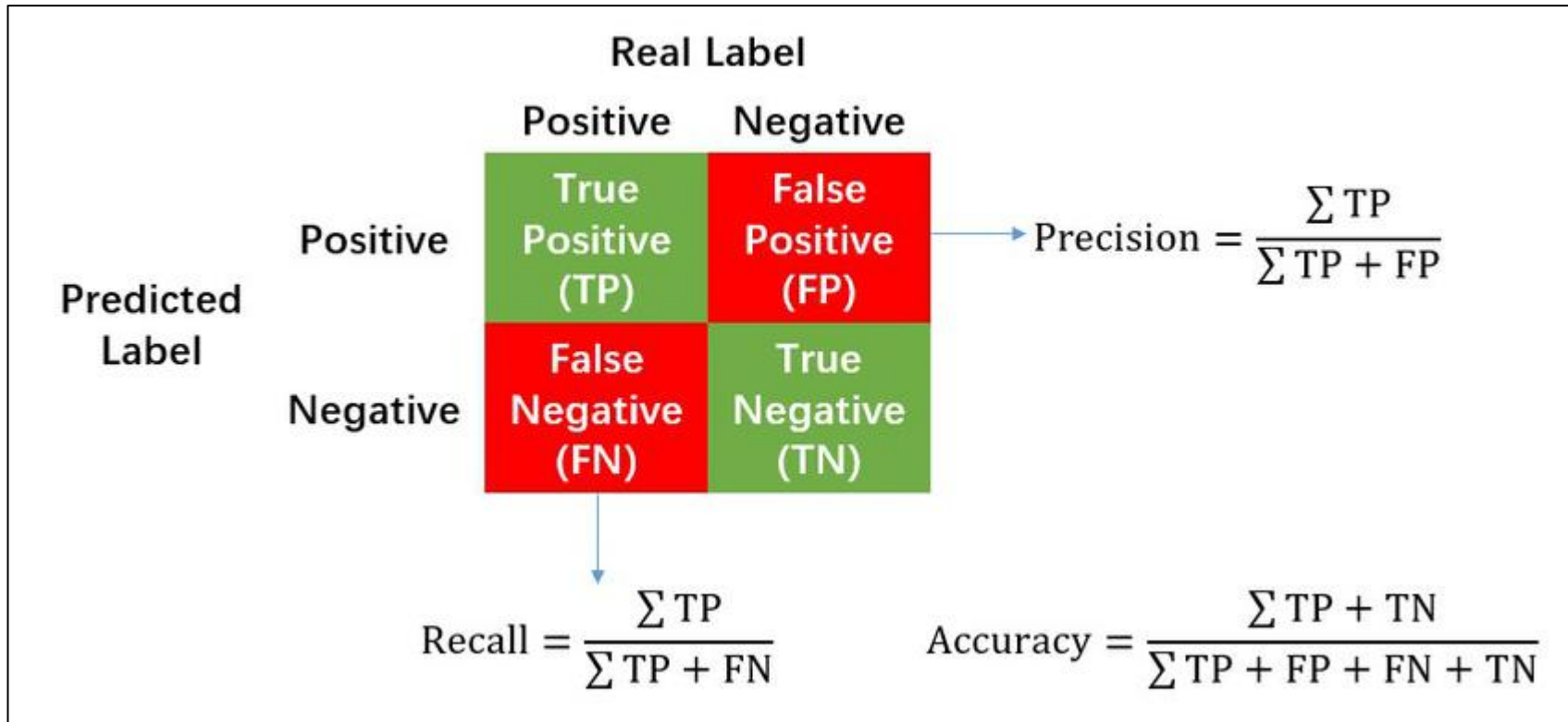
sepal_length	sepal_width	petal_length	petal_width	species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	Virginica
4.7	3.2	1.3	0.2	Virginica
4.6	3.1	1.5	0.2	setosa
5	3.6	1.4	0.2	Versicolor
5.4	3.9	1.7	0.4	setosa



Classification Metrics – Confusion Matrix

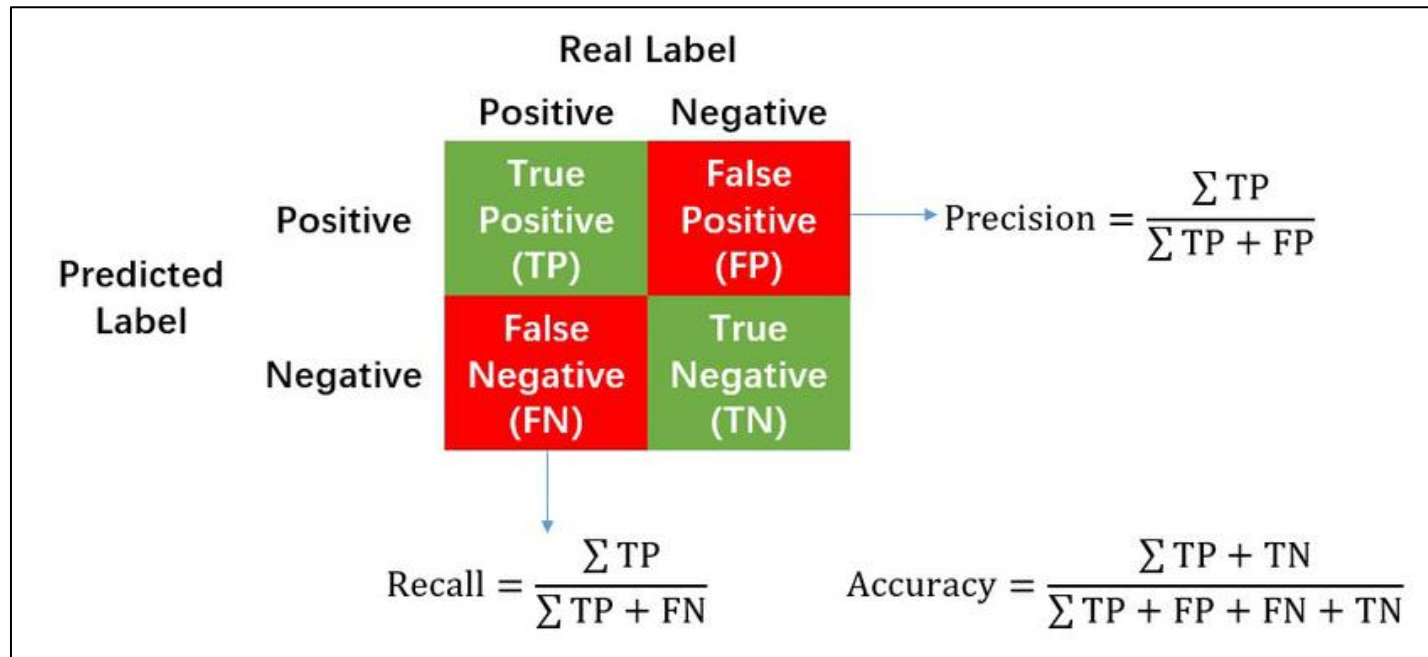


Precision, Recall, Accuracy



F1 Score

F1 Score is Harmonic mean of precision and recall



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

F-Beta Score

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

$$F_2 = (1 + 2^2) \cdot \frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} \cdot 2^2 + \textit{recall}}$$

$$F_2 = \frac{5 \cdot \textit{precision} \cdot \textit{recall}}{\textit{precision} \cdot 4 + \textit{recall}}$$

Classification Example where **precision** is important

- Email Spam and Not Spam classification
- Type 1 Error , Actual Mail is not spam but model predicted spam
- Because of this important mails can go into spam which is dangerous
- Hence Precision is more important than recall here

		Real Label		
		Positive	Negative	
Predicted Label	Positive	True Positive (TP)	False Positive (FP)	Precision = $\frac{\sum TP}{\sum TP + FP}$
	Negative	False Negative (FN)	True Negative (TN)	
		Recall = $\frac{\sum TP}{\sum TP + FN}$		Accuracy = $\frac{\sum TP + TN}{\sum TP + FP + FN + TN}$

Classification Example where Recall is more important

- Any Disease prediction eg. Checking whether a person is Diabetic or Not Diabetic
- Type 2 error – Person already Has Diabetes but model predicted non-Diabetic
- This can cause issues and worsen the health of the patient

		Real Label		
		Positive	Negative	
Predicted Label	Positive	True Positive (TP)	False Positive (FP)	Precision = $\frac{\sum TP}{\sum TP + FP}$
	Negative	False Negative (FN)	True Negative (TN)	
		Recall = $\frac{\sum TP}{\sum TP + FN}$		Accuracy = $\frac{\sum TP + TN}{\sum TP + FP + FN + TN}$

Thank you

FOR ANY QUERIES PING ME ON SKYPE