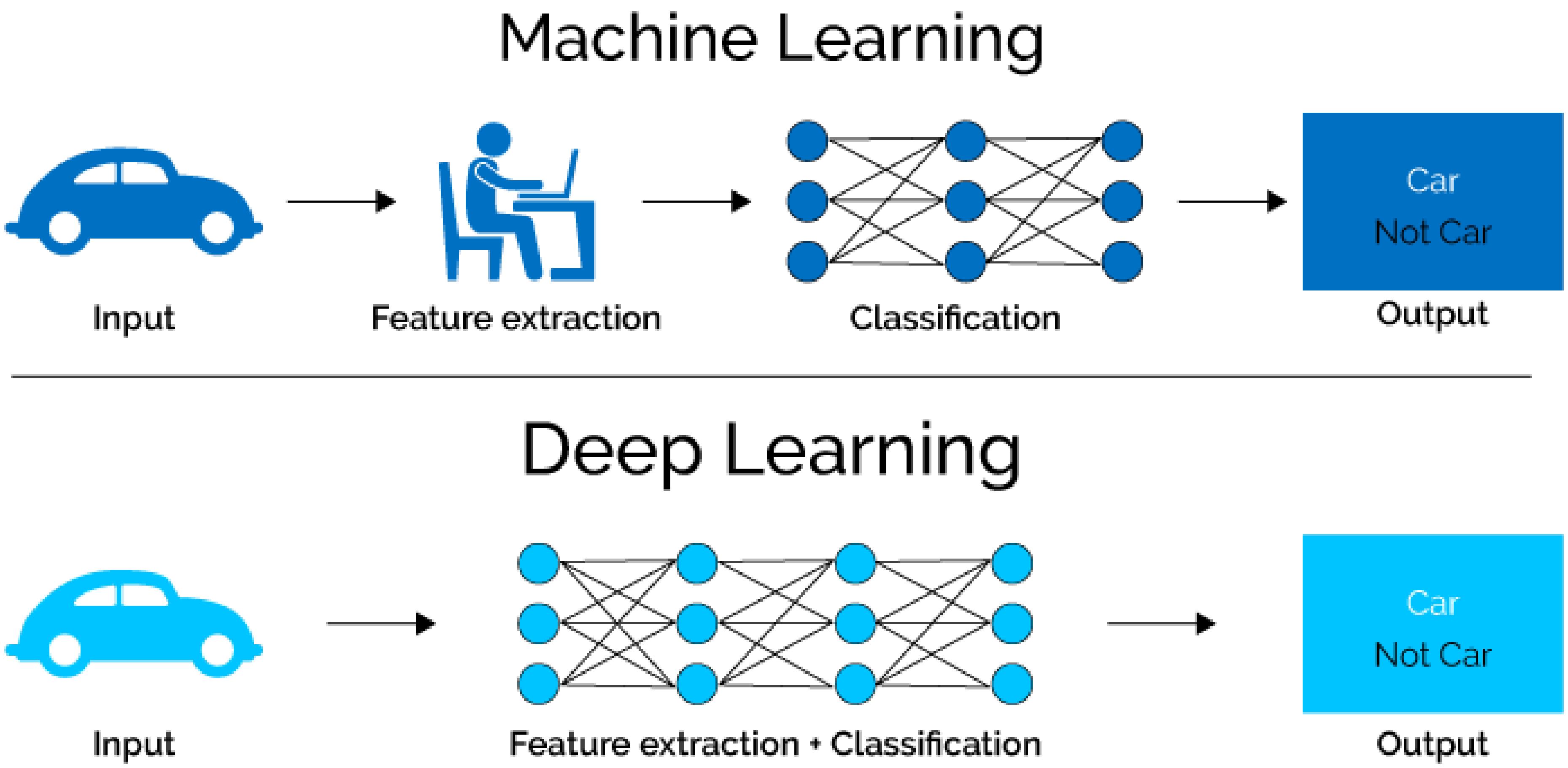
em lyon business school

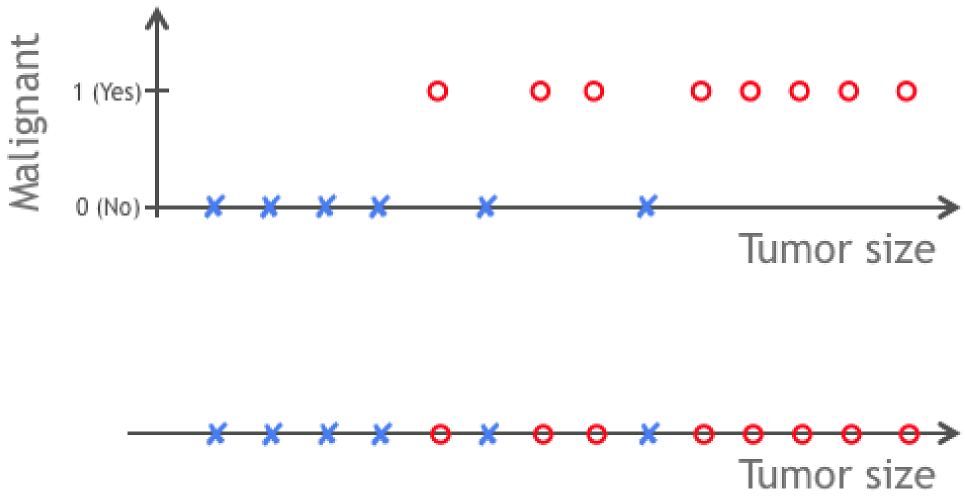
Session 7 – Classification

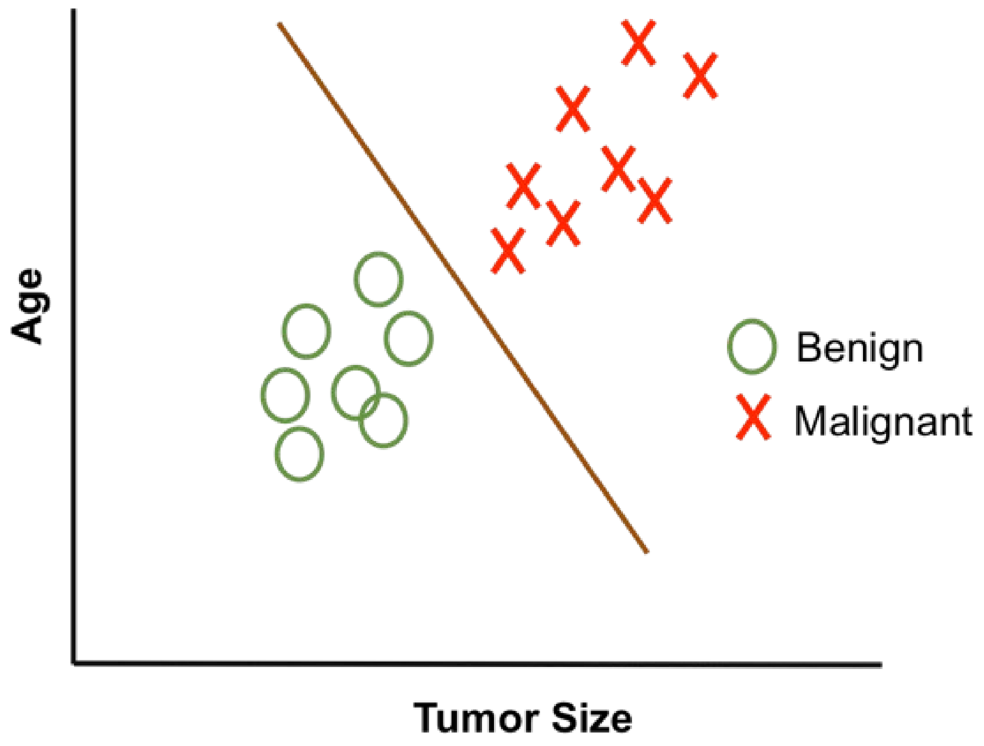
Application on image classification

A bit of Context

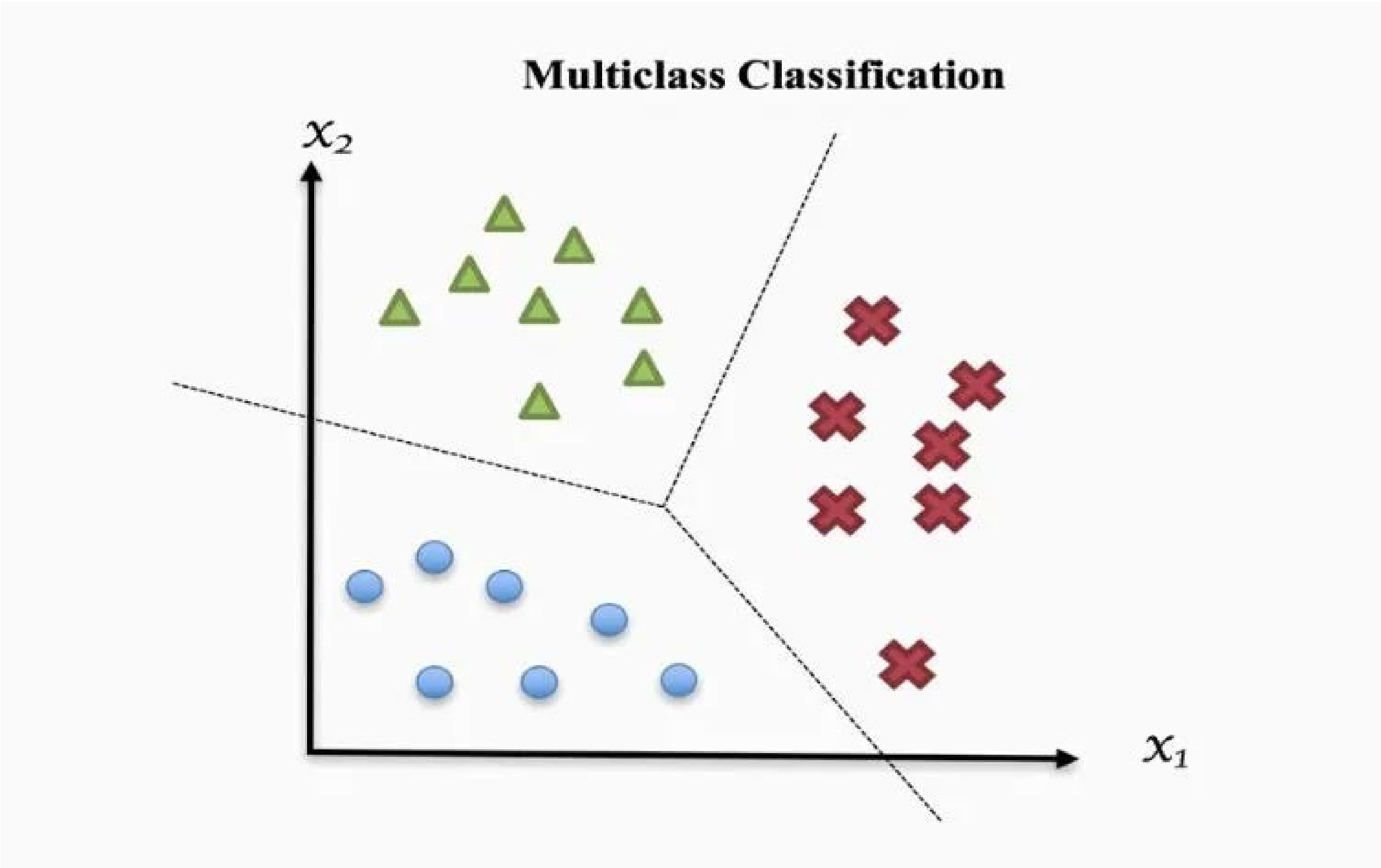


Binary classification

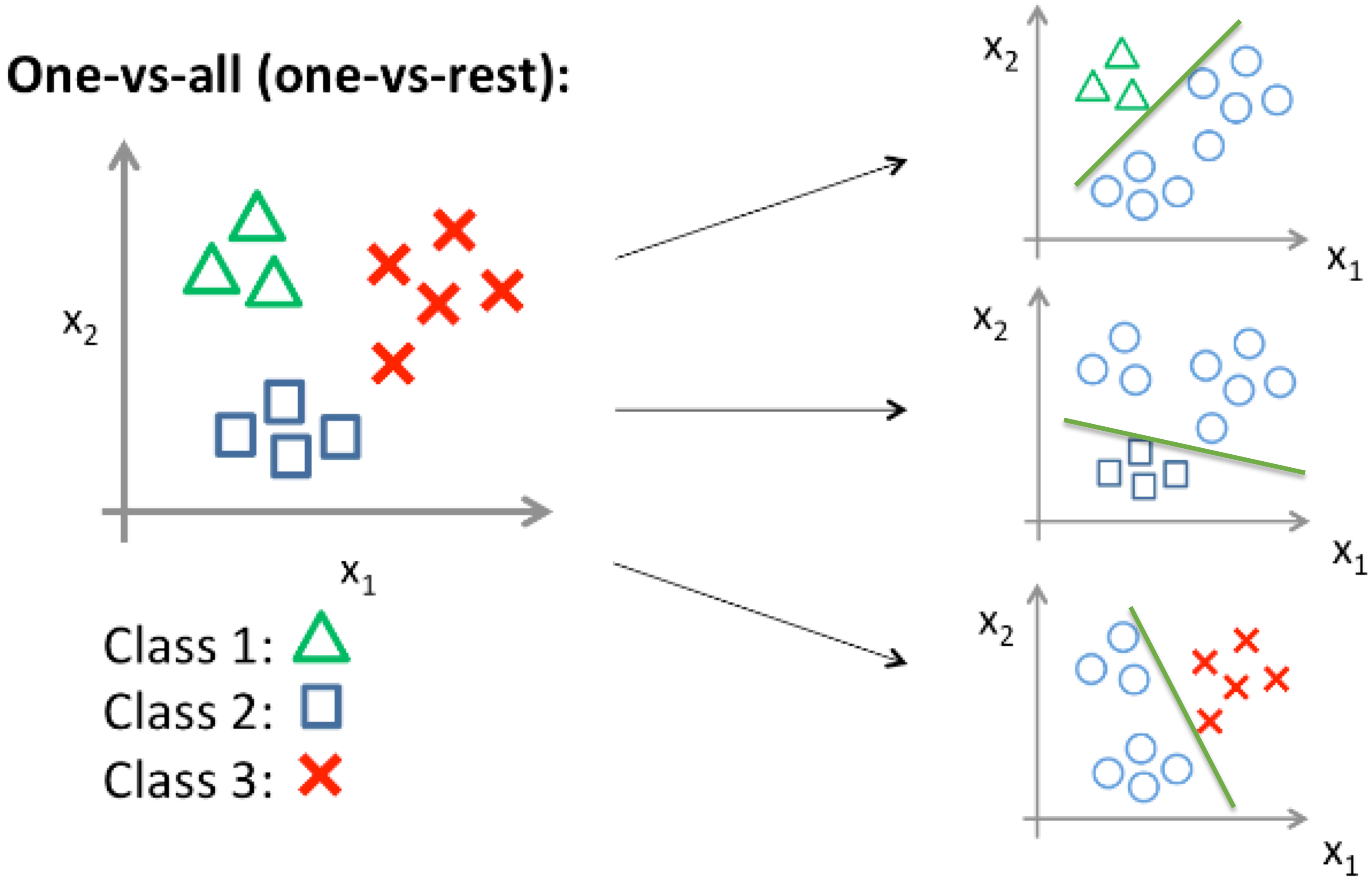




Multiclass Classification

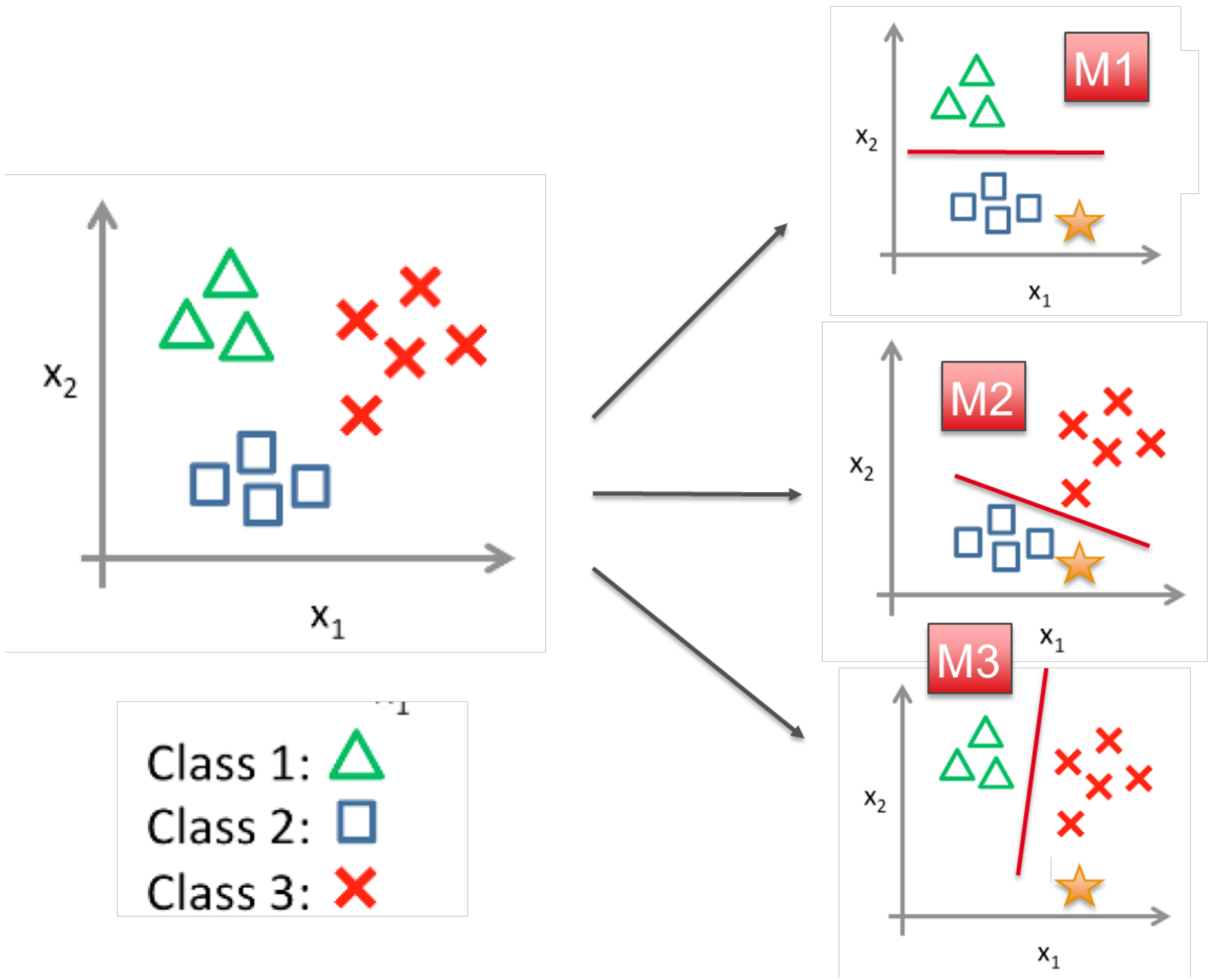


One vs all (OVA) or One vs Rest (OVR)



One vs one

Training



Prediction
Apply all classifiers to the new data point
Select class with majority vote

Model 1: □

Model 2: □

Model 3: ✕

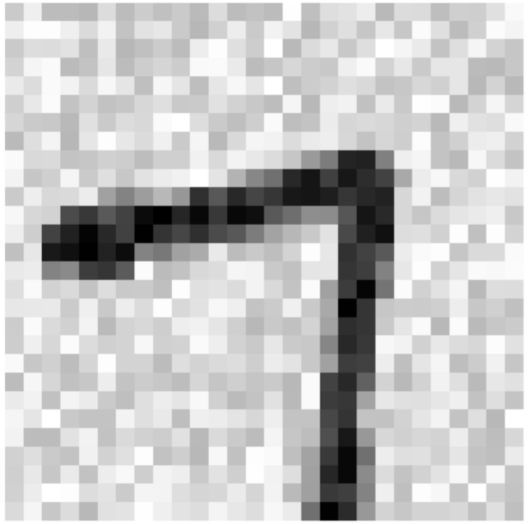
Majority Vote: □

Tie: Select class with highest Confidence

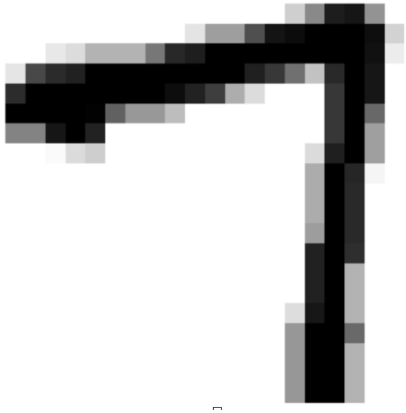
Multilabel Clssification

Multi-Output Classification

- Generalization of multi-label classification where each label can be multi-class.
- Example: System that removes noise from images



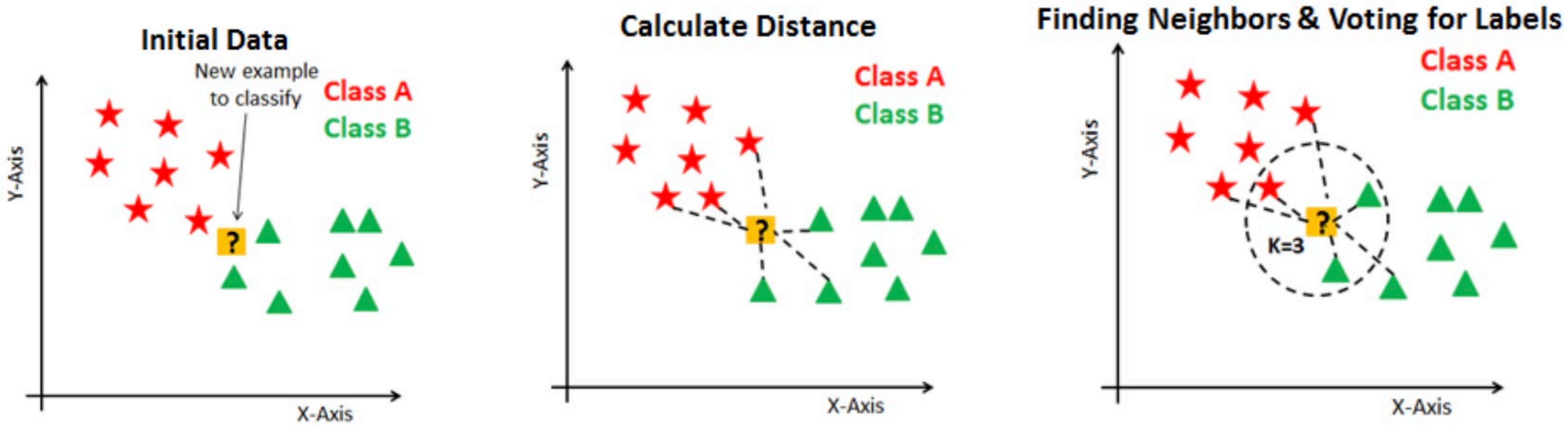
Input: Image + Noise



Output: Image without noise

Output: Multi-output
Multi-label: one label per pixel
Each label can have multiple values: Pixel intensity ranges
From 0 to 255
Output 1: Pixel intensity 1
:
:
Output n: Pixel intensity n

KNN Classifier



Decision Trees Classifier 1

Recommending Apps

Gender	Age	App
F	15	
F	25	
M	32	
F	40	
M	12	
M	14	

Between gender and age, which one seems more decisive for predicting what app the users will download?

Decision Trees Classifier 2

Recommending Apps

Gender	Age	App
F	15	
F	25	
M	32	
F	40	
M	12	
M	14	

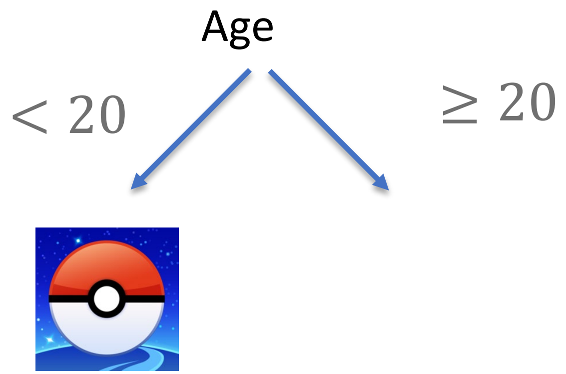
Between gender and age, which one seems more decisive for predicting what app the users will download?

Decision Trees Classifier 3

Recommending Apps

Gender	Age	App
F	15	
F	25	
M	32	
F	40	
M	12	
M	14	

Between gender and age, which one seems more decisive for predicting what app the users will download?

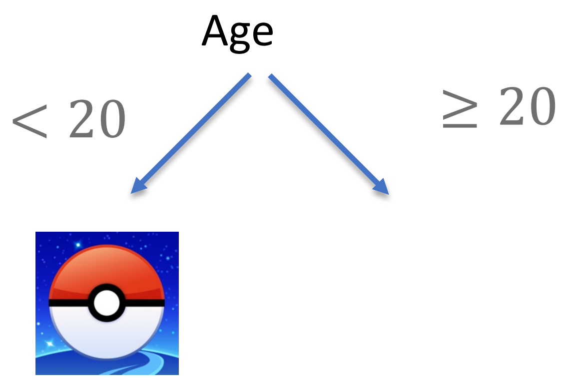


Decision Trees Classifier 4

Recommending Apps

Gender	Age	App
F	25	
M	32	
F	40	

Between gender and age, which one seems more decisive for predicting what app the users will download?

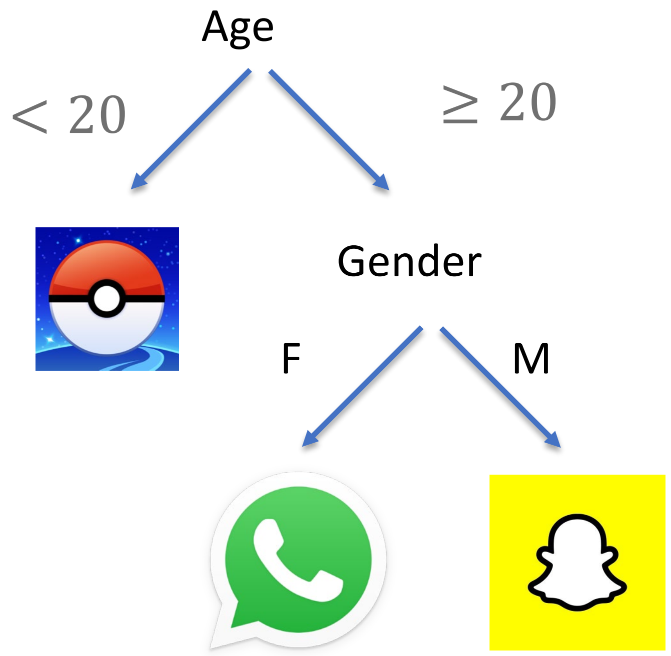


Decision Trees Classifier 5

Recommending Apps

Gender	Age	App
F	25	
M	32	
F	40	

Between gender and age, which one seems more decisive for predicting what app the users will download?



em lyon business school

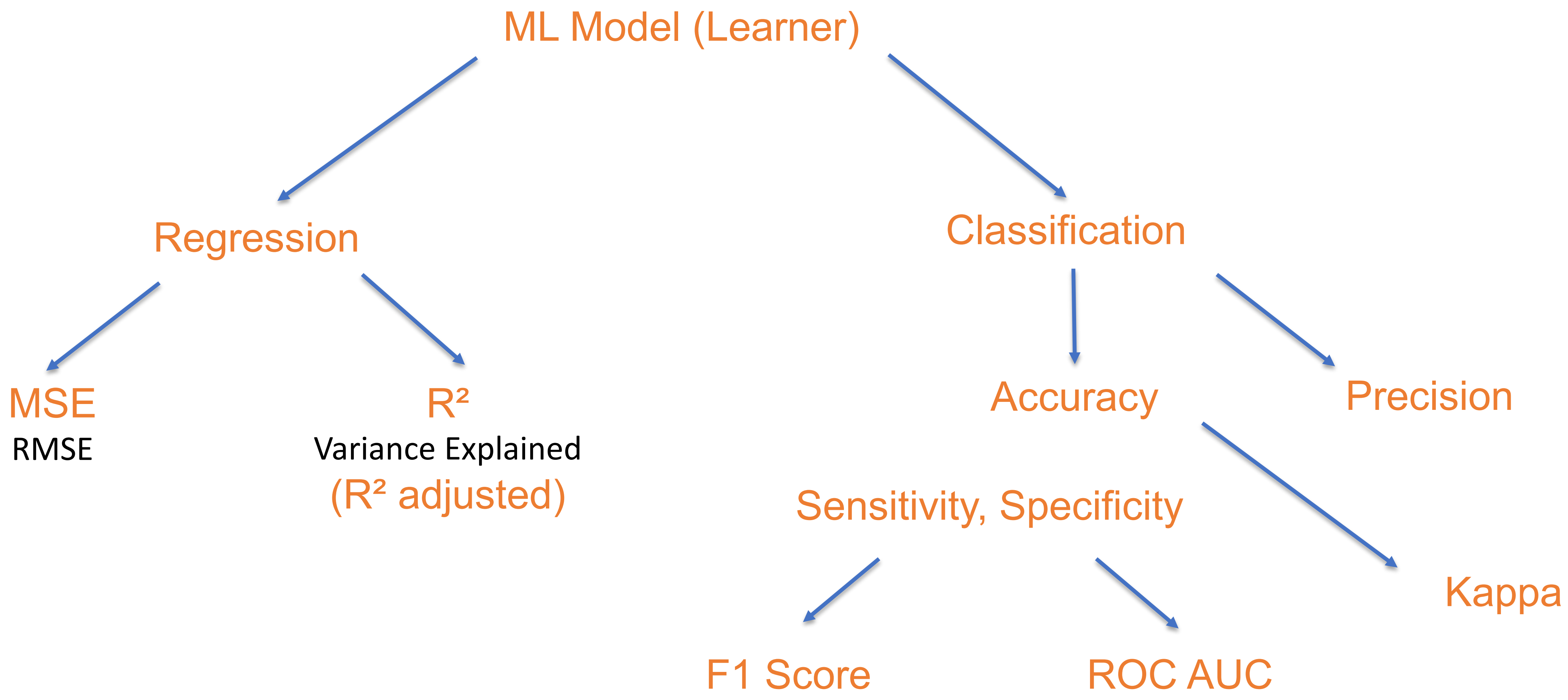
Other Classifiers

- **SVM Family**
  - Support Vector Machine Classifier can use different Kernels
    - Linear
    - Polynomial
    - Radial

- **Bagging Family**
  - Random Forest Classifier

- **Boosting Family**
  - Gradient Boosting Classifier (GBM)
  - XGBoost
  - LightGBM
  - Catboost

Model Evaluation



Confusion Matrix (Binary outcome or target)

How good is our model in predicting positive outcome (yes) ?

- Examine the classification table
  - The rows represent the actual classes of the sample
  - The columns represent the predicted classes – based on a cutoff probability value of 0.5
    - True negatives are the NO, predicted as NO (TN)
    - False positives are NO, predicted as YES (FP)
    - True positives are YES, predicted as YES (TP)
    - False negatives are YES, predicted as N (FN)

Confusion Matrix		Predicted		
		No	Yes	Total
Actual	No	TN	FP	Actual -
	Yes	FN	TP	Actual +
	Total	Predicted -	Predicted +	Sample size

$$Accuracy = \frac{TN + TP}{Sample\ size} \Leftrightarrow Error\ rate = 1 - Accuracy$$

$$Precision = \frac{TP}{Predicted +}$$

$$Sensitivity(Recall) = \frac{TP}{Actual +} = True\ positive\ rate\ (TPR)$$

$$Specificity = \frac{TN}{Actual -} = 1 - FPR\ (False\ positive\ rate)$$

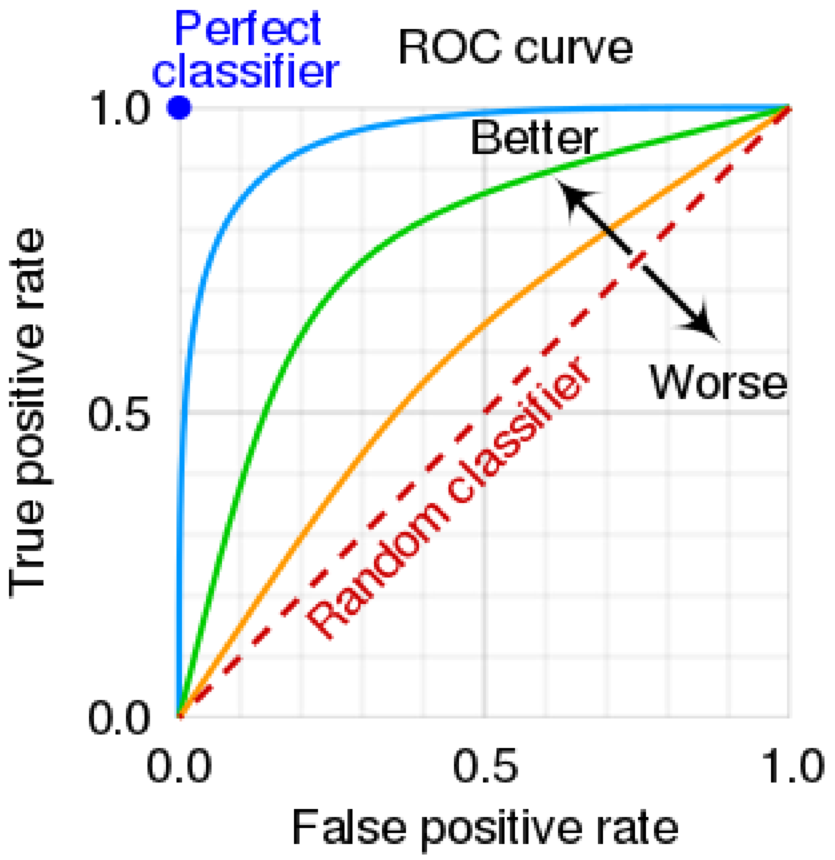
$$F_1\ Score = \frac{2 \times Recall\ \times Precision}{Recall + Precision}$$

The ROC AUC (AUROC)

- **Using these performance indicators usually makes several assumptions**
  - The 0.5 cutoff assumes approximately balanced datasets
  - The reliance on accuracy or error rate assumes equivalent costs of misclassification

- **Assume we have very unbalanced dataset**
  - Ex : Fraud detection : less than 5% of people are considered fraudulent
    - Any model predicting always people to be non fraudulent will have an accuracy of 95% or more
    - If 95% of the sample is negative, is it realistic to choose a probability of 0.5 to assign to the positive class ?
    - In a dataset of this type, the sensitivity is going to be very low while the specificity very high
    - Any indicator based on some threshold for class assignment is problematic for unbalanced dataset

- **One solution is to use a performance indicator that is threshold independent : the AUC (Area under the ROC Curve)**



Cohen's Kappa

- **Compares the actual (observed) accuracy to a random classifier's (expected) accuracy**

- **Better than the Accuracy**
  - More robust to imbalanced datasets
  - Applicable to multiclass classification
  - Remains threshold dependent
  - Fairly good classifier : $\boldsymbol{\kappa > 0.60}$

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

p_0 : probability of observed agreement
p_e: probability of expected (by chance) agreement