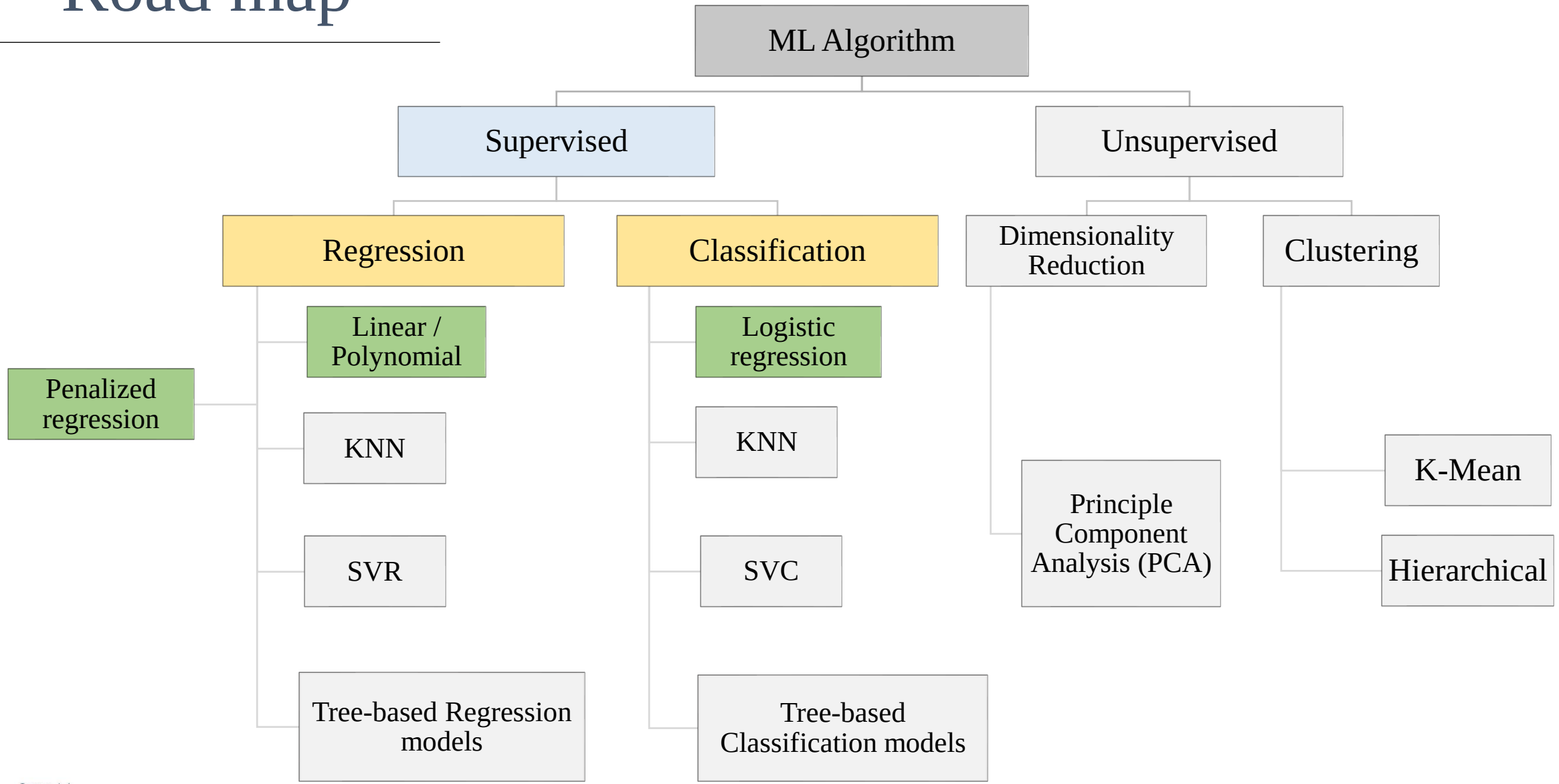
Class 12 – Classification metrics

Prof. Pedram Jahangiry

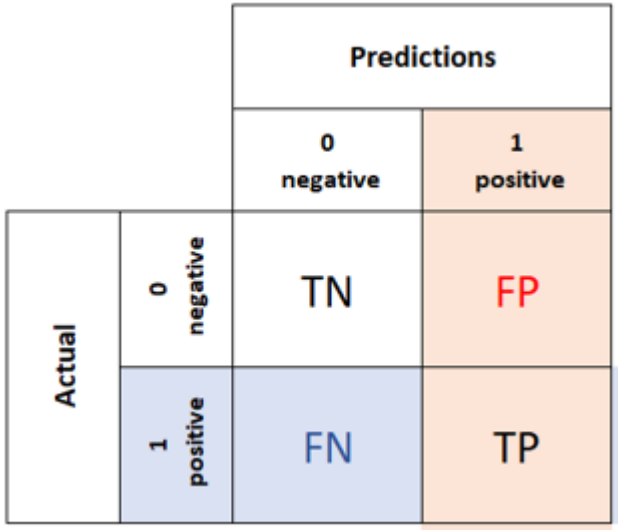
		Predictions	
		0 negative	1 positive
Actual	0 negative	TN	FP
	1 positive	FN	TP

Road map


ML Algorithm
Supervised
Unsupervised
Regression
Classification
Dimensionality Reduction
Clustering
Linear / Polynomial
Penalized regression
KNN
SVR
Tree-based Regression models
Logistic regression
KNN
SVC
Tree-based Classification models
Principle Component Analysis (PCA)
K-Mean
Hierarchical

Topics

- Classification performance metrics
  a) Accuracy
  b) Precision
  c) Recall
  d) F1 score
  e) MCC
  f) ROC and AUC



Confusion Matrix

		Predictions	
		0 negative	1 positive
Actual	0 negative	TN	FP*
	1 positive	FN**	TP

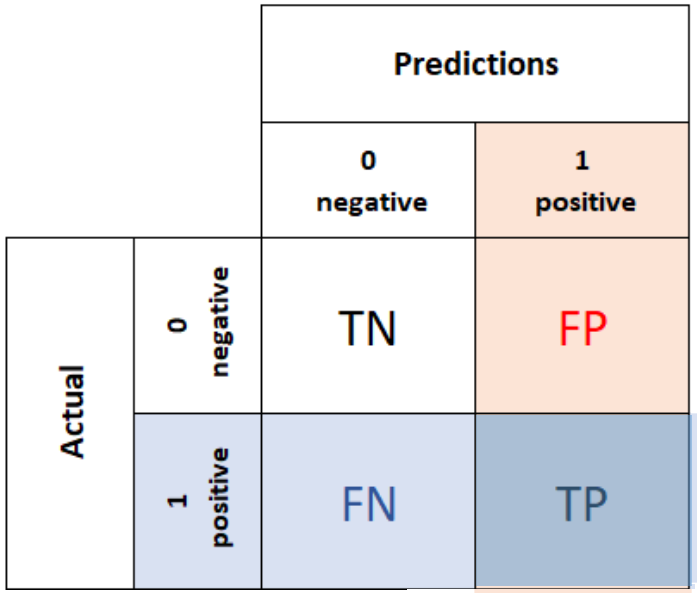
FP* Type I error

FN** Type II eror

		predicted class		
		class 1	class 2	class 3
actual class	class 1	True positives		
	class 2		True positives	
	class 3			True positives

Accuracy, Precision, Recall and F1score

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



While **recall** expresses the ability to find all relevant instances in a dataset, **precision** expresses the proportion of the data points our model says was relevant were actually relevant.

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

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

$$F1\ Score = 2 * \frac{PR}{P + R}$$

F1 uses the **harmonic** mean instead of a simple average because it punishes extreme values.

MCC (Matthews Correlation Coefficient)

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

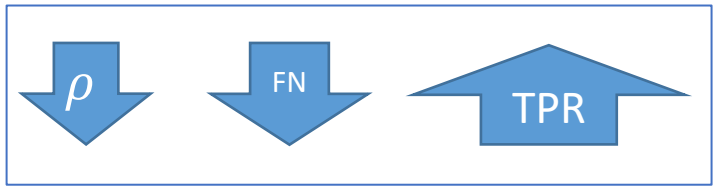
- Accuracy and error rates are misleading for imbalanced data sets.
- Precision, recall or even f1 score will not take into account the true negatives (TN)
- MCC is one of the most informative metrics for any binary classifier.

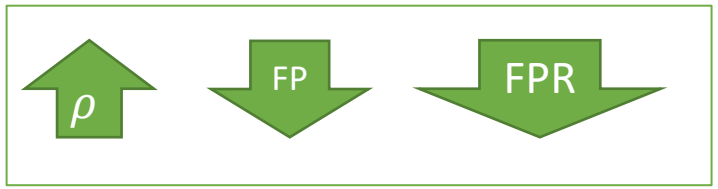
		Predictions	
		0 negative	1 positive
Actual	0 negative	TN	FP
	1 positive	FN	TP

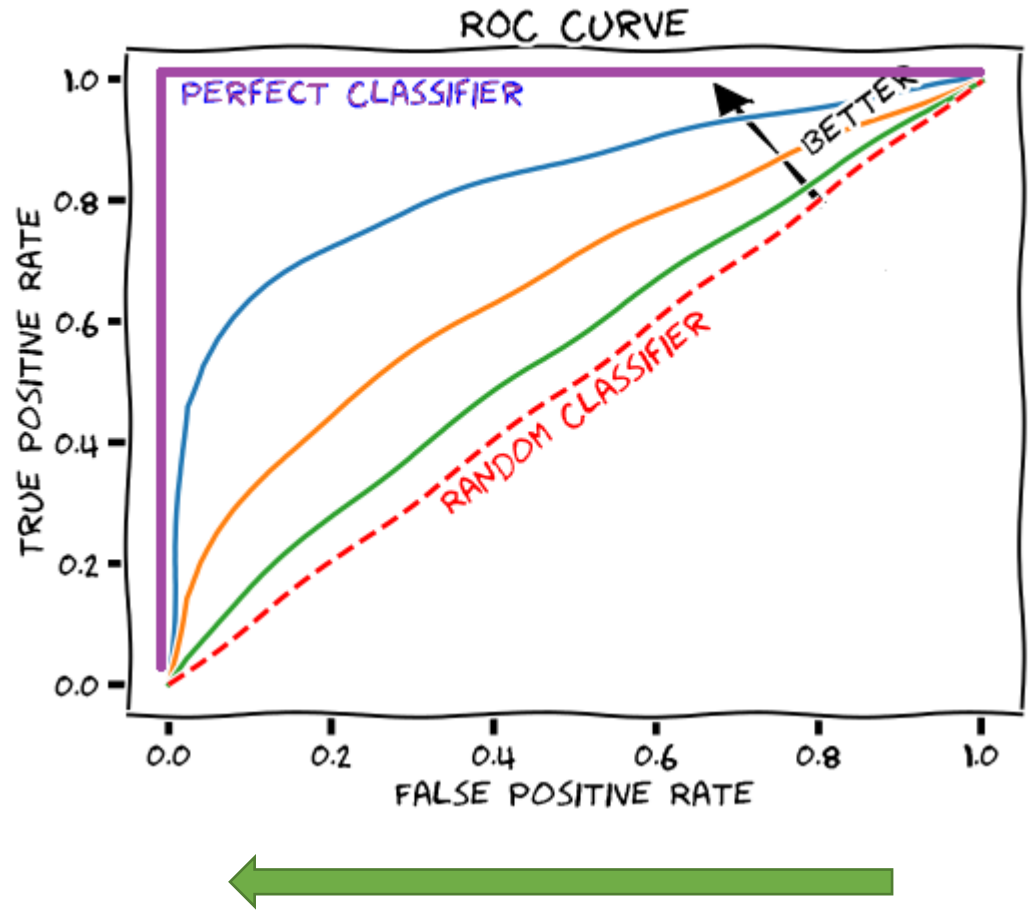
- MCC returns a value between -1 and +1.
  - +1 represents a perfect prediction,
  - 0 represents no better than a random prediction,
  - −1 indicates total misclassification

ROC (Receiver Operating Characteristic)

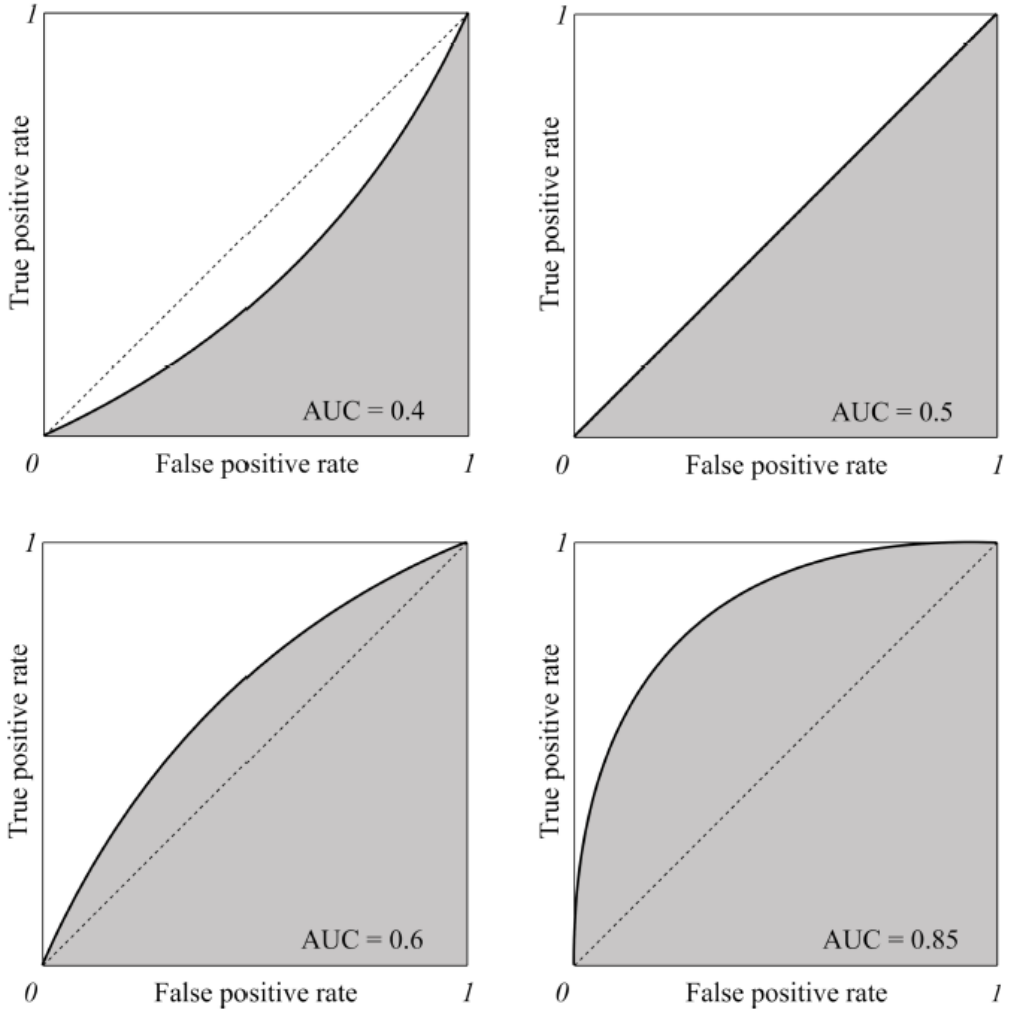
		Predictions: 0 negative	Predictions: 1 positive	
Actual	0 negative	TN	FP	$False\ Positive\ Rate = \frac{FP}{FP + TN}$
	1 positive	FN	TP	$True\ Positive\ Rate = \frac{TP}{TP + FN}$







AUC


1
True positive rate
AUC = 0.4
0
False positive rate
1
1
True positive rate
AUC = 0.5
0
False positive rate
1
1
True positive rate
AUC = 0.6
0
False positive rate
1
1
True positive rate
AUC = 0.85
0
False positive rate
1

Some other classification metrics

		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		


class
DISMISSED