

# Courbe ROC

Receiving Operating Characteristics  
Une autre manière d'évaluer un modèle de prédiction

Ricco RAKOTOMALALA



## Nécessité d'évaluer les modèles de prédiction

Évaluer les performances d'un modèle de prédiction est primordial

- 1** Pour savoir si le modèle est globalement significatif.  
Mon modèle traduit-il vraiment une causalité ?
- 2** Pour se donner une idée des performances en déploiement.  
Quelle sera la fiabilité (les coûts associés) lorsque j'utiliserai mon modèle ?
- 3** Pour comparer plusieurs modèles candidats.  
Lequel parmi plusieurs modèles sera le plus performant compte tenu de mes objectifs ?



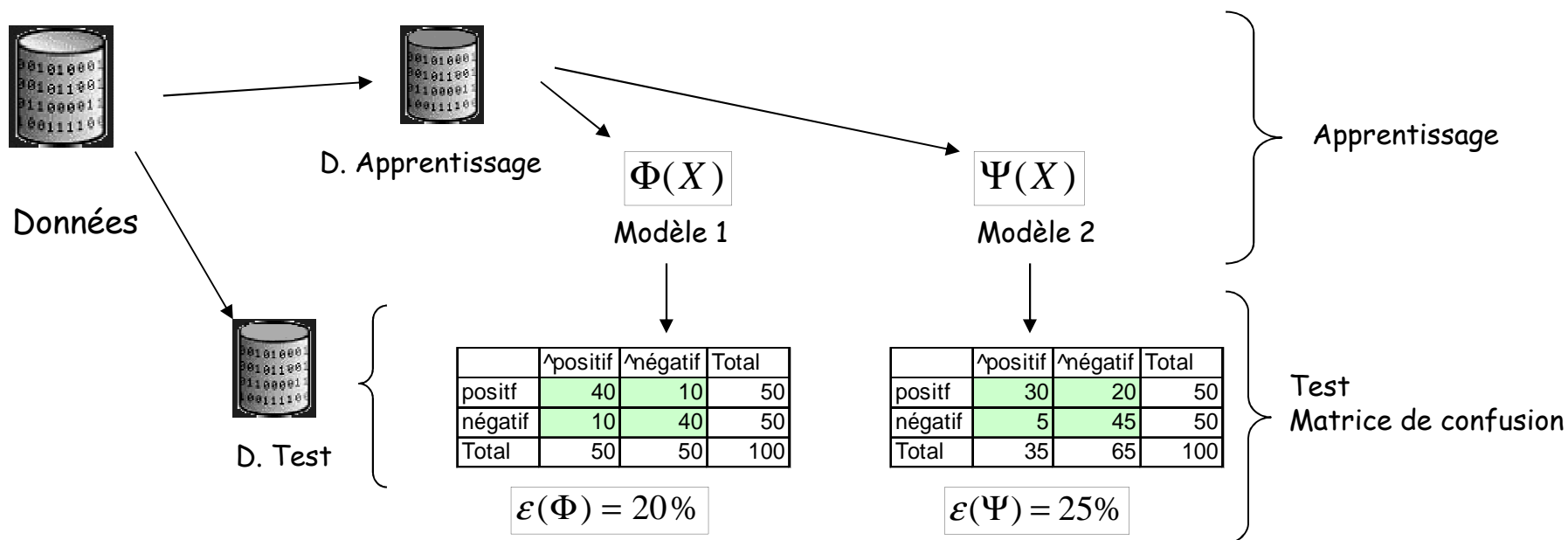
Le taux d'erreur semble être un indicateur synthétique pertinent, il indique (estime) la probabilité de mal classer un individu de la population.

Les autres indicateurs sont très intéressants également (sensibilité/rappel, précision/spécificité) mais obligent à surveiller plusieurs valeurs simultanément.



# Le taux d'erreur : un indicateur trop réducteur

## Schéma habituel d'évaluation des modèles



Conclusion : Modèle 1 serait meilleur que Modèle 2

Cette conclusion -- basée sur l'échantillon test -- suppose que la matrice de coût de mauvais classement est symétrique et unitaire



# Introduction d'une matrice de coût

## Coût de mauvaise affectation non-symétrique

	$\wedge$ positif	$\wedge$ négatif
positif	0	1
négatif	10	0

↙

	$\wedge$ positif	$\wedge$ négatif	Total
positif	40	10	50
négatif	10	40	50
Total	50	50	100

$$c(\Phi) = 1.1$$

↘

	$\wedge$ positif	$\wedge$ négatif	Total
positif	30	20	50
négatif	5	45	50
Total	35	65	100

$$c(\Psi) = 0.7$$

Conclusion : Modèle 2 serait meilleur que Modèle 1 dans ce cas ???

Les matrices de coûts sont souvent le fruit d'opportunités conjoncturelles.  
Faudrait-il tester toutes les matrices de coûts possibles pour comparer M1 et M2 ?



Peut-on bénéficier d'un dispositif qui permet de comparer globalement les modèles, indépendamment de la matrice de coût de mauvaise affectation ?



# Problème des distributions déséquilibrées

Lorsque les classes sont très déséquilibrées, la matrice de confusion et surtout le taux d'erreur donnent souvent une fausse idée de la qualité de l'apprentissage.

Ex. COIL 2000 - Challenge, détecter les personnes qui vont prendre une police d'assurance

## ANALYSE DISCRIMINANTE

Apprentissage

0.0627			
Confusion matrix			
	No	Yes	Sum
No	5435	39	5474
Yes	326	22	348
Sum	5761	61	5822

Test

0.0650			
Confusion matrix			
	No	Yes	Sum
No	3731	31	3762
Yes	229	9	238
Sum	3960	40	4000

Le classifieur par défaut (prédire systématiquement NO), propose un taux d'erreur de  $238 / 4000 = 0.0595$

Moralité : modéliser ne servirait à rien dès que les classes sont très déséquilibrées ?



Cette anomalie est liée au fait que nous voulons absolument que le modèle réalise un affectation (positif ou négatif). Or dans de nombreux domaines, ce qui nous intéresse avant tout, c'est de mesurer la propension à être positif ou négatif !



## Objectifs de la courbe ROC

La courbe ROC est un outil d'évaluation et de comparaison des modèles

- 1** Indépendant des matrices de coûts de mauvaise affectation  
Il permet de savoir si M1 sera toujours meilleur que M2 quelle que soit la matrice de coût
- 2** Opérationnel même dans le cas des distributions très déséquilibrées  
Sans les effets pervers de la matrice de confusion liés à la nécessité de réaliser une affectation
- 3** Résultats valables même si l'échantillon test n'est pas représentatif  
Tirage prospectif ou tirage rétrospectif : les indications fournies restent les mêmes
- 4** Un outil graphique qui permet de visualiser les performances  
Un seul coup d'œil doit permettre de voir le(s) modèle(s) susceptible(s) de nous intéresser
- 5** Un indicateur synthétique associé  
Aisément interprétable

Sa portée va largement au-delà des interprétations (indicateurs)  
issues de l'analyse de la matrice de confusion.



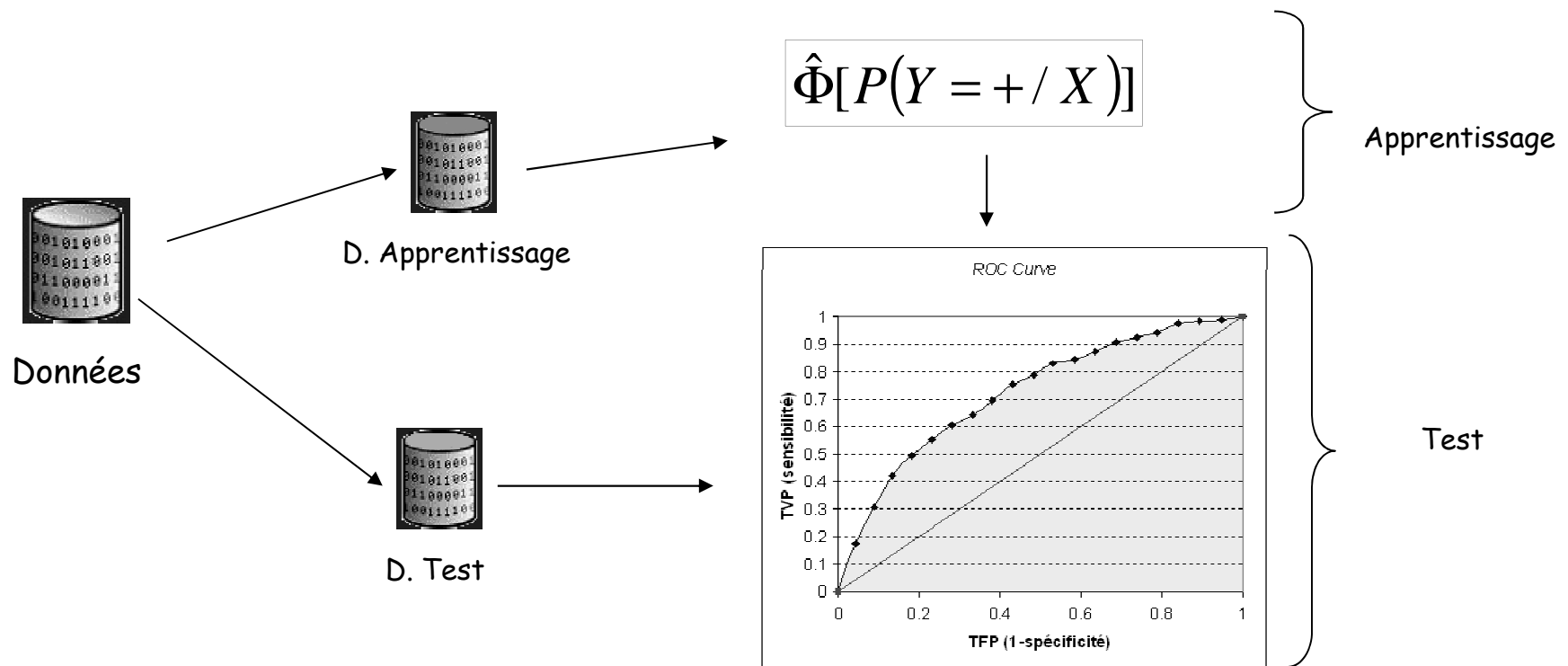
## Cadre de l'utilisation de la courbe ROC

Nous sommes dans un problème à 2 classes

En fait dans tous les cas où nous avons la possibilité de définir la modalité (+) de la variable à prédire

Le modèle de prédiction fournit  $P(Y=+/X)$

Ou toute grandeur proportionnelle à  $P(Y=+/X)$  qui permettra de classer les observations



La similitude avec la courbe LIFT est tentante, mais la lecture et l'interprétation de la courbe ROC est totalement différente



# Principe de la courbe ROC

## Matrice de confusion

	^positif	^négatif
positif	VP	FN
négatif	FP	VN

TVP = Rappel = Sensibilité = VP/Positifs

TFP = 1 - Spécificité = FP/Négatifs

### Principe de la courbe ROC

$P(Y=+/X) \geq P(Y=-/X)$  équivaut à une règle d'affectation  $P(Y=+/X) \geq 0.5$  (seuil = 0.5)

→ Cette règle d'affectation fournit une matrice de confusion MC1, et donc 2 indicateurs TVP1 et TFP1

Si nous choisissons un autre seuil (0.6 par ex.), nous obtiendrons MC2 et donc TVP2 et TF2  
Etc... MCi, TVPi, TFPi



L'idée de la courbe ROC est de faire varier le « seuil » de 1 à 0 et, pour chaque cas, calculer le TVP et le TFP que l'on reporte dans un graphique : en abscisse le TFP, en ordonnée le TVP.





# Construction de la courbe ROC (1/2)

Classer les données  
selon un score décroissant

Individu	Score (+)	Classe
1	1	+
2	0.95	+
3	0.9	+
4	0.85	-
5	0.8	+
6	0.75	-
7	0.7	-
8	0.65	+
9	0.6	-
10	0.55	-
11	0.5	-
12	0.45	+
13	0.4	-
14	0.35	-
15	0.3	-
16	0.25	-
17	0.2	-
18	0.15	-
19	0.1	-
20	0.05	-

Positifs = 6  
Négatifs = 14

Seuil = 1

	^positif	^négatif	Total
positif	1	5	6
négatif	0	14	14
Total	1	19	20

TVP = 1/6 = 0.2 ; TFP = 0/14 = 0

Seuil = 0.95

	^positif	^négatif	Total
positif	2	4	6
négatif	0	14	14
Total	2	18	20

TVP = 2/6 = 0.33 ; TFP = 0/14 = 0

Seuil = 0.9

	^positif	^négatif	Total
positif	3	3	6
négatif	0	14	14
Total	3	17	20

TVP = 3/6 = 0.5 ; TFP = 0/14 = 0

Seuil = 0.85

	^positif	^négatif	Total
positif	3	3	6
négatif	1	13	14
Total	4	16	20

TVP = 3/6 = 0.5 ; TFP = 1/14 = 0.07

Seuil = 0

	^positif	^négatif	Total
positif	6	0	6
négatif	14	0	14
Total	20	0	20

TVP = 6/6 = 1 ; TFP = 14/14 = 1



## Construction de la courbe ROC (2/2)

Mettre en relation

TFP (abscisse) et TVP (ordonnée)

Individu	Score (+)	Classe	TFP	TVP
			0	0.000
1	1	+	0.000	0.167
2	0.95	+	0.000	0.333
3	0.9	+	0.000	0.500
4	0.85	-	0.071	0.500
5	0.8	+	0.071	0.667
6	0.75	-	0.143	0.667
7	0.7	-	0.214	0.667
8	0.65	+	0.214	0.833
9	0.6	-	0.286	0.833
10	0.55	-	0.357	0.833
11	0.5	-	0.429	0.833
12	0.45	+	0.429	1.000
13	0.4	-	0.500	1.000
14	0.35	-	0.571	1.000
15	0.3	-	0.643	1.000
16	0.25	-	0.714	1.000
17	0.2	-	0.786	1.000
18	0.15	-	0.857	1.000
19	0.1	-	0.929	1.000
20	0.05	-	1.000	1.000

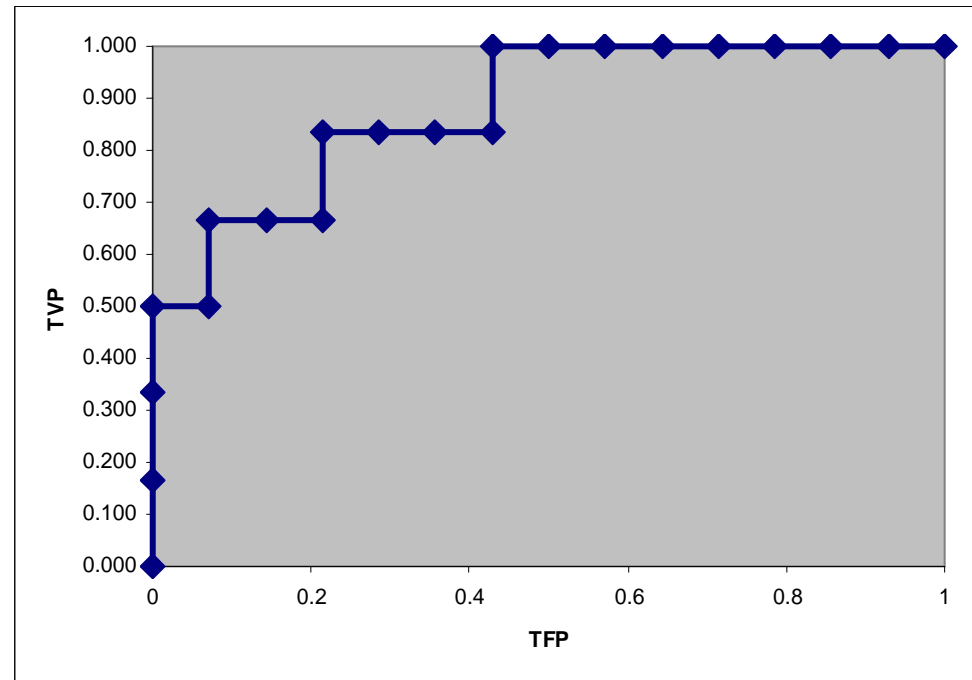


### Calcul pratique

$TFP(i) = \text{Nombre de négatifs parmi les « i » premiers} / (\text{nombre total des négatifs})$

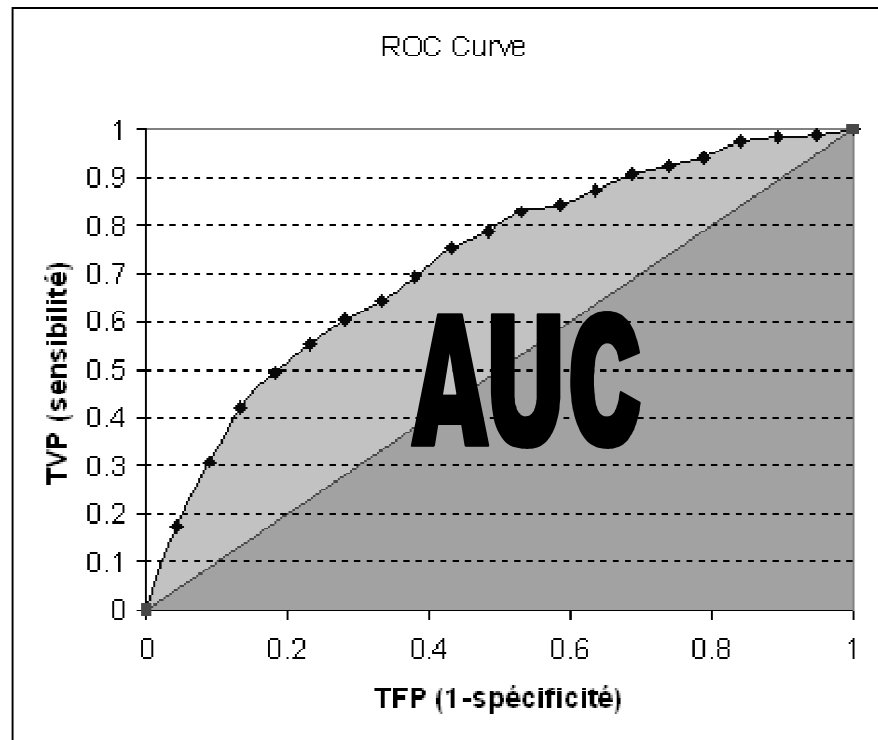
$TVP(i) = \text{Nombre de positifs parmi les « i » premiers} / (\text{nombre total des positifs})$

### Courbe ROC



Interprétation : AUC, l'aire sous la courbe

AUC indique la probabilité pour que la fonction SCORE place un positif devant un négatif (dans le meilleur des cas  $AUC = 1$ )



Si SCORE classe au hasard les individus (c.-à-d. le modèle de prédiction ne sert à rien),  $AUC = 0.5$

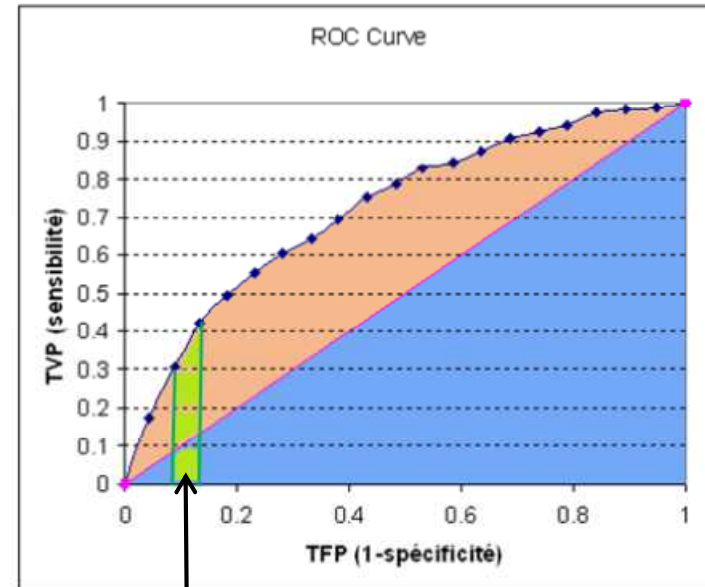
→ Symbolisée par la diagonale principale dans le graphique



# AUC – Calcul pratique 1 – Intégration avec la méthode des trapèzes

Dérivée directement de la définition : surface = intégrale

Individu	Score (+)	Classe	TFP	TVP	Largeur	Hauteur	Surface
			0	0.000			
1	1	+	0.000	0.167	0.000	0.083	0.000
2	0.95	+	0.000	0.333	0.000	0.250	0.000
3	0.9	+	0.000	0.500	0.000	0.417	0.000
4	0.85	-	0.071	0.500	0.071	0.500	0.036
5	0.8	+	0.071	0.667	0.000	0.583	0.000
6	0.75	-	0.143	0.667	0.071	0.667	0.048
7	0.7	-	0.214	0.667	0.071	0.667	0.048
8	0.65	+	0.214	0.833	0.000	0.750	0.000
9	0.6	-	0.286	0.833	0.071	0.833	0.060
10	0.55	-	0.357	0.833	0.071	0.833	0.060
11	0.5	-	0.429	0.833	0.071	0.833	0.060
12	0.45	+	0.429	1.000	0.000	0.917	0.000
13	0.4	-	0.500	1.000	0.071	1.000	0.071
14	0.35	-	0.571	1.000	0.071	1.000	0.071
15	0.3	-	0.643	1.000	0.071	1.000	0.071
16	0.25	-	0.714	1.000	0.071	1.000	0.071
17	0.2	-	0.786	1.000	0.071	1.000	0.071
18	0.15	-	0.857	1.000	0.071	1.000	0.071
19	0.1	-	0.929	1.000	0.071	1.000	0.071
20	0.05	-	1.000	1.000	0.071	1.000	0.071
<b>AUC</b>							<b>0.881</b>



$$s_i = (TFP_i - TFP_{i-1}) \times \frac{TVP_i + TVP_{i-1}}{2}$$

Surface d'un trapèze

$$AUC = \sum_i s_i$$

AUC = SOMME (Surface des trapèzes)



## AUC – Calcul pratique 2 – La statistique de Mann-Whitney

Individu	Score (+)	Classe	Rangs	Rangs +
1	1	+	20	20
2	0.95	+	19	19
3	0.9	+	18	18
4	0.85	-	17	0
5	0.8	+	16	16
6	0.75	-	15	0
7	0.7	-	14	0
8	0.65	+	13	13
9	0.6	-	12	0
10	0.55	-	11	0
11	0.5	-	10	0
12	0.45	+	9	9
13	0.4	-	8	0
14	0.35	-	7	0
15	0.3	-	6	0
16	0.25	-	5	0
17	0.2	-	4	0
18	0.15	-	3	0
19	0.1	-	2	0
20	0.05	-	1	0
Somme (Rang +)				95
U+				74
<b>AUC</b>				<b>0.881</b>

Test de Mann-Whitney : montrer que deux distributions sont différentes (décalées).

Statistique basée sur les rangs.

Dans notre contexte, montrer que les « + » présentent en moyenne des scores plus élevés que les « - ».

On peut en dériver un test statistique.

Somme des rangs des « + »

$$S_+ = \sum_{i: y_i=+} r_i = 20 + 19 + 18 + 16 + 13 + 9 = 95$$

Statistique de Mann-Whitney

$$U_+ = S_+ - \frac{n_+(n_+ + 1)}{2} = 95 - \frac{6 \times 7}{2} = 74$$

AUC

$$AUC = \frac{U_+}{n_+ \times n_-} = \frac{74}{6 \times 14} = 0.881$$



## AUC – Calcul pratique 3 – Dénombrer les inversions

Individu	Score (+)	Classe	Nb de "-" devant un "+"
1	1	+	0
2	0.95	+	0
3	0.9	+	0
4	0.85	-	0
5	0.8	+	1
6	0.75	-	0
7	0.7	-	0
8	0.65	+	3
9	0.6	-	0
10	0.55	-	0
11	0.5	-	0
12	0.45	+	6
13	0.4	-	0
14	0.35	-	0
15	0.3	-	0
16	0.25	-	0
17	0.2	-	0
18	0.15	-	0
19	0.1	-	0
20	0.05	-	0
<b>Swaps</b>			<b>10</b>
<b>AUC</b>			<b>0.881</b>

Trier les individus selon un score décroissant.  
 Pour chaque « + », compter le nombre de « - » qui le précède.  
 Dans notre contexte, on souhaite que les scores élevés soient attribués aux « + » en priorité c.-à-d. les « + » sont peu précédés de « - ».

Swaps : somme de l'indicateur ci-dessus

$$Swaps = \sum_{i: y_i = +} c_i = 0 + 0 + 0 + 1 + 3 + 6 = 10$$

AUC

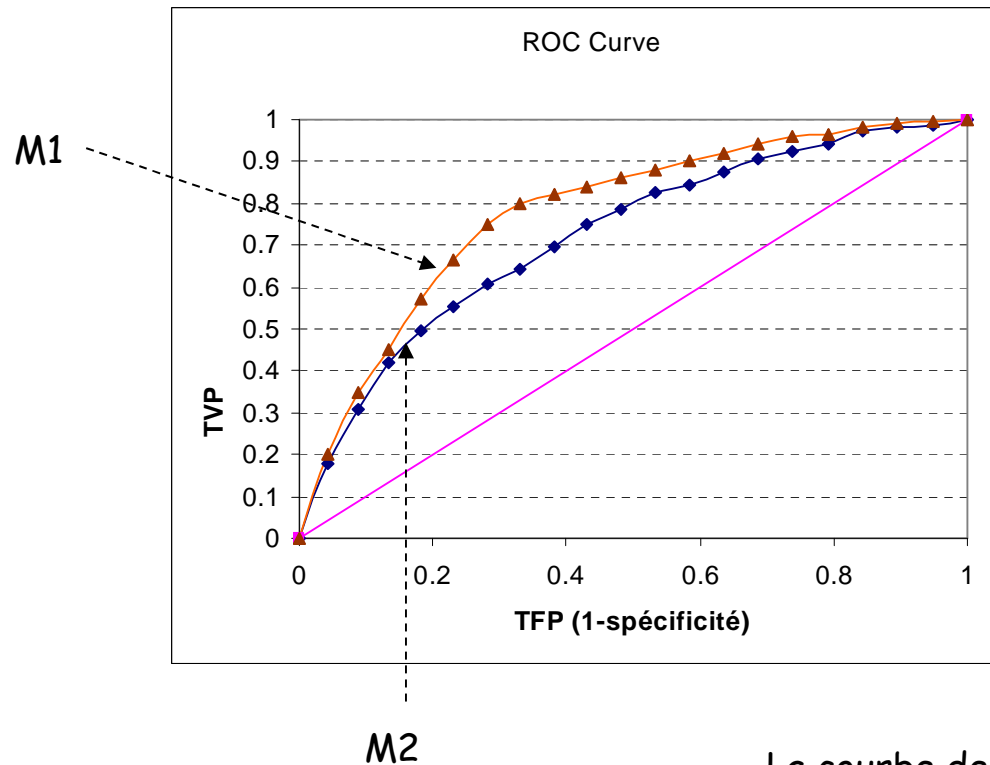
$$AUC = 1 - \frac{Swaps}{n_+ \times n_-} = 1 - \frac{10}{6 \times 14} = 0.881$$



## Interprétation : dominance



Comment montrer que M1 sera toujours meilleur que M2, quelle que soit la matrice de coût de mauvaise affectation utilisée ?



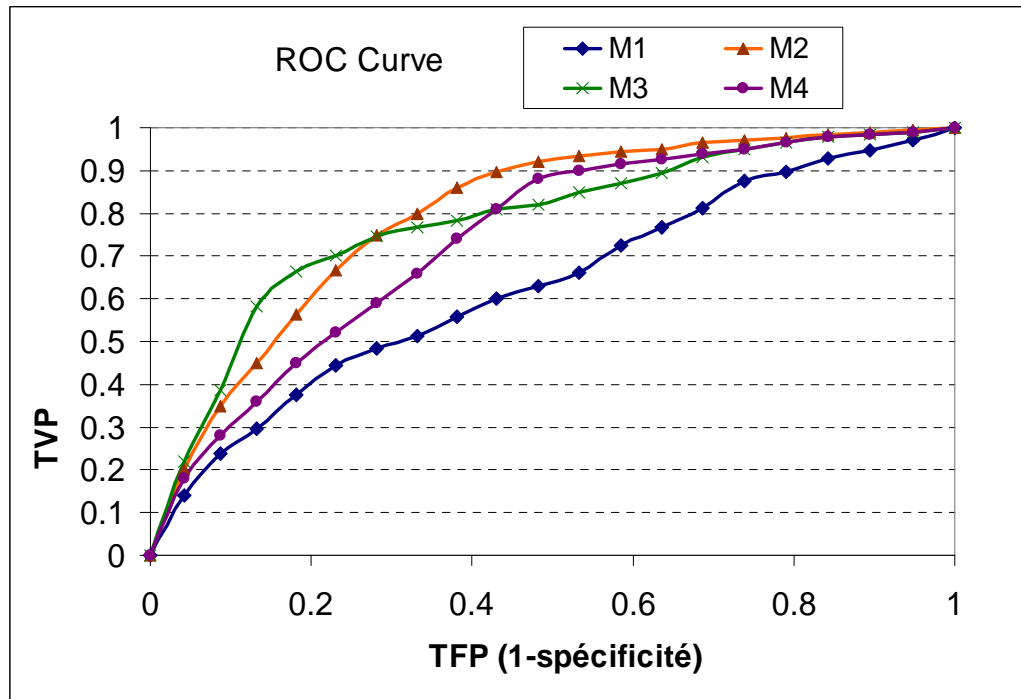
La courbe de M1 est toujours « au-dessus » de celle de M2 : il ne peut pas exister de situation (matrice de coût de mauvais classement) où M2 serait un meilleur modèle de prédiction.



# Enveloppe convexe : sélection de modèles



Parmi un ensemble de modèles candidats, comment éliminer d'office ceux qui ne seront pas intéressants ?



## Enveloppe convexe

Elle est formée par les courbes qui, à un moment ou à un autre, n'ont aucune courbe « au-dessus » d'elles.

Les courbes situées sur cette enveloppe correspondent aux modèles qui sont potentiellement les plus performantes pour une matrice de coût donnée.

Les modèles qui ne participent jamais à cette enveloppe peuvent être éliminés.

Dans notre exemple, l'enveloppe convexe est formée par les courbes de M3 et M2.



- » M1 est dominé par tous les modèles, il peut être éliminé.
- » M4 peut être meilleur que M3 dans certains cas, mais dans ces cas là, il sera moins bon que M2 : M4 peut être éliminé.





# Conclusion

Dans de nombreuses applications, la courbe ROC fournit des informations plus intéressantes sur la qualité de l'apprentissage que le simple taux d'erreur.



» C'est surtout vrai lorsque les classes sont très déséquilibrées, et lorsque le coût de mauvaise affectation est susceptible de modifications.



» Il faut néanmoins que l'on ait une classe cible (positive) clairement identifiée et que la méthode d'apprentissage puisse fournir un SCORE proportionnel à  $P(Y=+/X)$ .

