

Fouille de données

NGUYỄN Thị Minh Huyền

Email: huyenntm@vnu.edu.vn

VNU University of Science, Hanoi

Plan

1. Mesures de qualité

2. Courbe ROC

3. Courbe de lift

Plan

1. Mesures de qualité

2. Courbe ROC

3. Courbe de lift

Plan

1. Mesures de qualité

2. Courbe ROC

3. Courbe de lift

Plan

1. Mesures de qualité

2. Courbe ROC

3. Courbe de lift

Classification

- Classification binaire : positive ou négative.
- Classification probabiliste : $f(x) \in [0, 1]$, seuil t .
 $f(x) \geq t \Rightarrow x$ positive, sinon négative
 \Rightarrow classification binaire en fonction de t .

Mesures de qualité

- Ensemble de test : P cas positifs, N cas négatifs.
- Valeurs $TP(t)$ (*true positive*), $FP(t)$, $TN(t)$, $TF(t)$
- $TPrate = TP/P$ (**Rappel**), $FPrate = FP/N$,
 $YRate = (TP + FP)/(P + N)$
- **Précision** = $TP/(TP + FP)$, **Exactitude** (*Accuracy*) =
 $(TP + TN)/(P + N)$
- F-mesure =
 $Précision * Rappel / (\alpha Précision + (1 - \alpha) Rappel)$,
 $(\alpha \in [0, 1])$
 $\alpha = 0.5 \Rightarrow$
 $F1 = 2 * Précision * Rappel / (Précision + Rappel) .$

Plan

1. Mesures de qualité

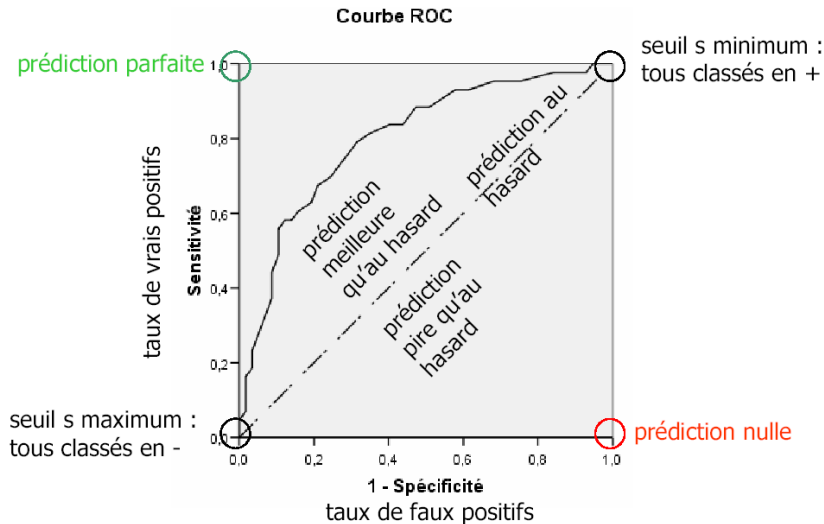
2. Courbe ROC

3. Courbe de lift

Courbe ROC (*receiver operating characteristics*)

- Pour chaque fonction f , la courbe ROC est définie par $x = FPrate(t)$, $y = TPrate(t)$ en variant le seuil t .
- Mesure AUC (*Area Under Curve*) : surface sous la courbe ROC = 1 pour un modèle idéal, = 0,5 pour un modèle aléatoire
⇒ un bon modèle a une valeur AUC entre 0,5 et 1.

Courbe ROC



Plan

1. Mesures de qualité

2. Courbe ROC

3. Courbe de lift

Courbe de lift

- Pour chaque fonction f , la courbe de lift est définie par $x = Yrate(t)$, $y = TP(t)$ en variant le seuil t .
- Mesure AUC (*Area Under Curve*) : surface sous la courbe lift = P pour un modèle idéal, = $P/2$ pour un modèle aléatoire
⇒ un bon modèle a une valeur AUC entre $P/2$ et P .

Courbe de lift : exemple

