

Courbes ROC et Precision Recall

Ghazi Bel Mufti

ghazi.belmufti@gmail.com

ESSAI-2 / DATA MINING

Plan

- 1 introduction
- 2 Sensibilité et spécificité
- 3 Cas du scoring
- 4 Precision et Recall

Introduction

On cherche à évaluer la qualité d'un modèle.

La mesure de qualité la plus simple est le **taux de bien classés** :

$$\tau = \frac{\text{nombre de cas bien classés}}{\text{nombre total de cas}}$$

Cependant le taux de bien classés n'aide pas à une prise de décision basée sur l'évaluation des risques d'erreur qui peuvent dépendre :

- 1 du contexte ;
- 2 de l'utilisateur des résultats de l'analyse.

Insuffisance de la matrice de confusion : trois exemples

① Des affectations différentes pour des probabilités très proches

un individu avec $\pi(x_i) = 0.495$ sera désigné "négatif", un autre avec $\pi(x_{i'}) = 0.505$ sera désigné "positif". Pourtant, si l'on se réfère aux probabilités, ils sont finalement assez proches.

② Cas d'une classe d'effectif faible

Exemple d'une maladie rare. Supposons que 5% d'une population soit atteinte d'une maladie. Un test qui ne détecte jamais cette maladie, a un taux de bien classés égal à 95%.

③ Une même valeur de précision pour des risques différents

Deux règles de décision peuvent avoir la même précision et des taux différents en ce qui concernent les décisions correctes et incorrectes.

Les utilités de ces règles peuvent alors différer selon les contextes qui se présentent.

Sensibility et Specificity I

| Appart. | Décision | T+ | T- |
|------------|----------|------|------|
| | | | |
| ω_1 | | (TP) | (FN) |
| ω_2 | | (FP) | (TN) |

Le taux de bien classés ou (i.e. **Accuracy**) est un indicateur trop simple :

$$\text{Sensitivity} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}}$$

$$\text{Specificity} = \frac{\text{True Negative (TN)}}{\text{True Negative (TN)} + \text{False Positive (FP)}}$$

Sensitivity et Specificity II

REMARQUES :

1. True Positive + False Negative = Positive cases = $\# \omega_1$
2. True Negative + False Positive = Negative cases = $\# \omega_2$
3. Sensibility et Specificity peuvent s'interpréter comme des indicateurs de précision restreints, respectivement, aux ensembles des cas positifs et négatifs.

Notations :

TPR = pourcentage de vrais positifs = sensibility

TNR = pourcentage de vrais négatifs = specificity

FPR = pourcentage de faux positifs = $\frac{\# \text{ faux positifs}}{\# \text{ cas négatifs}}$

FNR = pourcentage de faux négatifs = $\frac{\# \text{ faux négatifs}}{\# \text{ cas positifs}}$

REMARQUE :

$$\text{TPR} + \text{FNR} = \text{TNR} + \text{FPR} = 1$$

Donc, deux de ces rapports sont suffisants pour connaître tous les pourcentages de décision.

Règle R1

| Décision Appart. | | T+ | T- | Total |
|---------------------|------------|----------|----------|-------|
| | | | | |
| | ω_1 | 140 (TP) | 60 (FN) | 200 |
| | ω_2 | 100 (FP) | 900 (TN) | 1000 |
| | Total | 240 | 960 | |

$TPR = 0.7$; $FNR = 0.3$; $FPR = 0.1$; $TNR = 0.9$; $Prec. = 0.87$

$\mathbb{P}(\omega_1) = 0.17$; $\mathbb{P}(\omega_2) = 0.83$

Règle R2

| Décision Appart. | | T+ | T- | Total |
|---------------------|------------|---------|----------|-------|
| | | | | |
| | ω_1 | 80 (TP) | 120 (FN) | 200 |
| | ω_2 | 40 (FP) | 960 (TN) | 1000 |
| | Total | 120 | 1080 | |

1) Pour la règle R2, calculer TPR, FNR, FPR, TNR et Precision.

2) Comparer les deux règles R1 et R2

Cas du scoring

Cas du scoring

Score S : à tout individu (objets, etc ...) est associé un nombre $S(x)$ permettant l'affectation de l'individu x à l'une des classes.

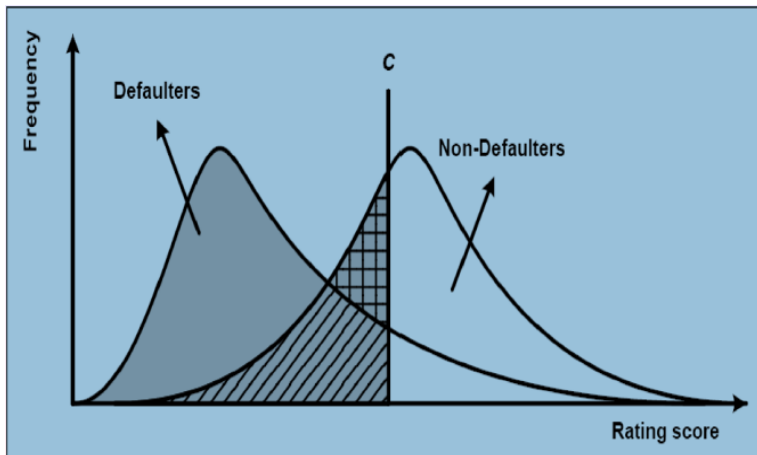
Cas du scoring avec deux classes

Un seul score suffit : par exemple, la probabilité a posteriori de l'appartenance à l'une des classes dans le cas d'une analyse discriminante bayésienne.

REMARQUE :

Il est fréquent que les distributions du score S conditionnellement aux classes ω_1 et ω_2 se chevauchent.

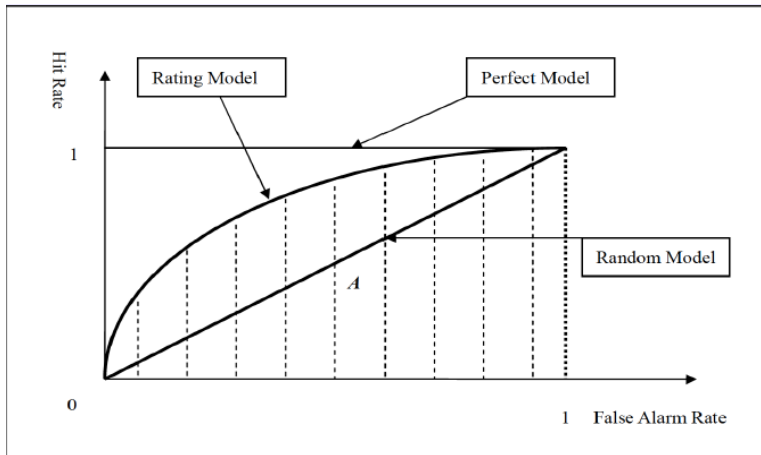
Distributions conditionnelles de S



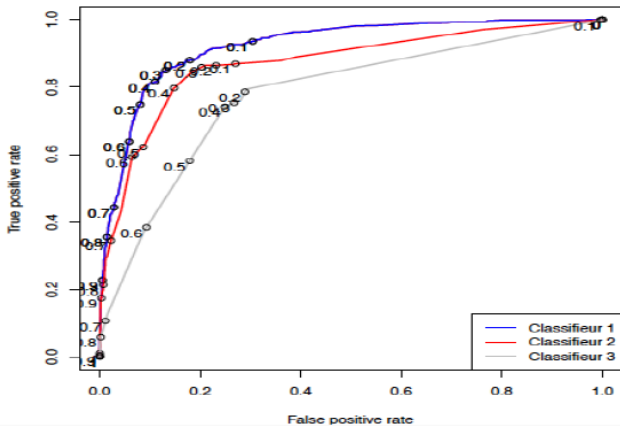
Remarques

- ① Augmenter le seuil s de décision
 - ⇒ décisions de vrais et faux positifs ↘
 - ⇒ décisions de vrais et faux négatifs ↗
 - ↳ Recherche d'un compromis
- ② On veut considérer et comparer plusieurs seuils de décision de façon à prendre en compte la nature subjective de la décision.
- ③ Comme deux pourcentages bien choisis, par ex. TPR et FPR, suffisent pour déterminer les deux autres, il suffit de calculer la valeur de ces deux pourcentages pour différents seuils de décision s .
- ④ On a vu que si s diminue, alors FPR et TPR, i.e. (1 - specificity) et sensibility augmentent.
Courbe ROC : graphe des paires (FPR, TPR) lorsque s varie.

Courbe ROC



Exemple de courbes ROC



Evaluation

- 1 Un score S est meilleur qu'un autre score S' si la courbe de S est toujours au dessus de celle de S' .
- 2 La surface sous la courbe ROC, notée **AUC** (aire sous la courbe), mesure la performance du score. On a :

$$\text{Coeff. de Gini} = 2AUC - 1$$

avec :

$$AUC = \mathbb{P}(X_1 > X_2),$$

où X_1 et X_2 suivent la loi du score S sachant ω_1 et ω_2 , respect., et où cette probabilité est calculée en supposant que l'on tire au hasard et indépendemment une observation de ω_1 et une observation de ω_2 :

L'air sous la courbe ROC n'est autre que la probabilité que le score d'un individu x soit supérieur à celui d'un individu y , si x est tiré aléatoirement parmi les positifs et y est tiré parmi les négatifs.

Precision et Recall I

| Appart. \ Décision | T+ | T- |
|--------------------|------|------|
| | | |
| ω_1 | (TP) | (FN) |
| ω_2 | (FP) | (TN) |

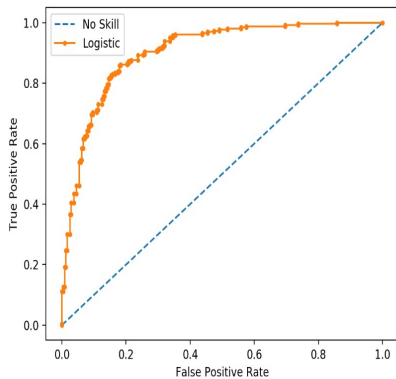
$$\text{Recall (i.e. Sensitivity)} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}}$$

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}}$$

REMARQUES :

1. True Positive + False Positive = Positive predicted cases = # T₊
2. Precision : pourc. de vrais positifs parmi les prédits positifs.
3. Precision et Recall se concentrent sur la classe positive et ne se soucient pas des vrais négatifs.

Cas de données équilibrées I



ROC Curve of a Logistic Regression Model and a No Skill Classifier

Figure – ROC AUC 0.903

Cas de données équilibrées II

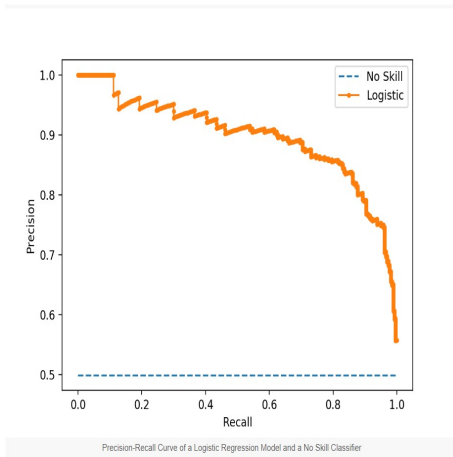
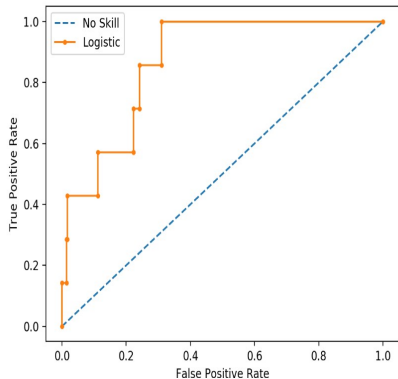


Figure – PR AUC : 0.898

Cas de données déséquilibrées I

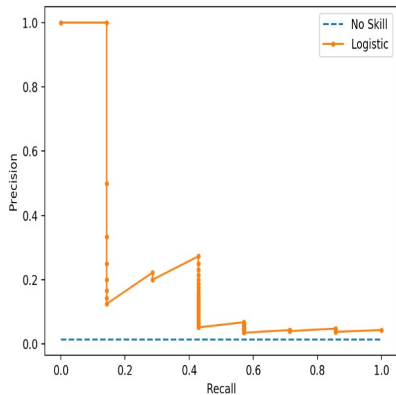
Dataset: Class0=985, Class1=15



Plot of ROC Curve for Logistic Regression on Imbalanced Classification Dataset

Figure – ROC AUC : 0.869

Cas de données déséquilibrées II



Plot of Precision-Recall Curve for Logistic Regression on Imbalanced Classification Dataset

Figure – PR AUC : 0.228

Cas de données déséquilibrées III

”To explain why the ROC and PR curves tell a different story, recall that the PR curve focuses on the minority class, whereas the ROC curve covers both classes.”

RÉFÉRENCES :

① L'AUC Precision-Recall

<https://kobia.fr/classification-metrics-auc-pr/>

② ROC Curves and Precision-Recall Curves for Imbalanced Classification

<https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-imbalanced-classification/>