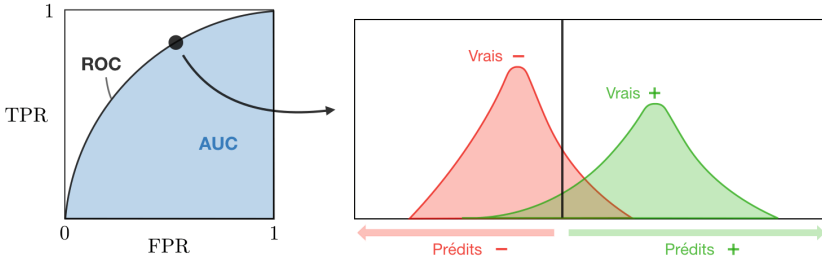


■ **AUC** – L'aire sous la courbe ROC, aussi notée AUC (de l'anglais *Area Under the Curve*) ou AUROC (de l'anglais *Area Under the ROC*), est l'aire sous la courbe ROC comme le montre la figure suivante :



4.2 Indicateurs dans le contexte de la régression

□ **Indicateurs de base** – Étant donné un modèle de régression f , les indicateurs suivants sont communément utilisés pour évaluer la performance d'un modèle :

Somme des carrés totale	Somme des carrés expliquée	Somme des carrés résiduelle
$SS_{\text{tot}} = \sum_{i=1}^m (y_i - \bar{y})^2$	$SS_{\text{reg}} = \sum_{i=1}^m (f(x_i) - \bar{y})^2$	$SS_{\text{res}} = \sum_{i=1}^m (y_i - f(x_i))^2$

□ **Coefficient de détermination** – Le coefficient de détermination, souvent noté R^2 ou r^2 , donne une mesure sur la qualité du modèle et est tel que :

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

□ **Indicateurs principaux** – Les indicateurs suivants sont communément utilisés pour évaluer la performance des modèles de régression, en prenant en compte le nombre de variables n qu'ils prennent en considération :

Cp de Mallow	AIC	BIC	R^2 ajusté
$\frac{SS_{\text{res}} + 2(n+1)\hat{\sigma}^2}{m}$	$2[(n+2) - \log(L)]$	$\log(m)(n+2) - 2\log(L)$	$1 - \frac{(1-R^2)(m-1)}{m-n-1}$

où L est la vraisemblance et $\hat{\sigma}^2$ est une estimation de la variance associée à chaque réponse.

4.3 Sélection de modèle

□ **Vocabulaire** – Lors de la sélection d'un modèle, on divise les données en 3 différentes parties comme suit :