Machine Learning
TD1

1. Considérons 2 variables binaires X_1 et X_2. Le tableau 1 décrit les lois jointes $P(X_1, X_2)$ et conditionnelles $P(Y = 1|X_1, X_2)$:

X_1	X_2	$P(X_1, X_2)$	$P(Y = 1\|X_1, X_2)$
0	0	0.3	0.6
0	1	0.3	0.2
1	0	0.2	0.4
1	1	0.2	0.8

Table 1 – Distribution $P(Y, X_1, X_2)$

— Quel est le classifieur "Bayes optimal" $f^\star(x_1, x_2)$ de y ? Que vaut son erreur de généralisation ?
— Construire un classifieur naïf de Bayes et calculer son erreur de généralisation. Comparez-le à $f^\star$.
— Construire un arbre de décision de niveau 1 (avec le gain de gini) et calculer son erreur de généralisation. Comparez-le à $f^\star$.
— Construire un échantillon de 10 exemples par un procédé de tirage aléatoire que vous explicitez. Quelle est sa log vraisemblance d'après le vrai modèle ? Construire un classifieur naïf de Bayes et un arbre de décision de niveau 1. Comparez les performances avec les précédentes. Conclure.

2. Quelle est l'entropie d'unve variable aléatoire gaussienne ?

3. On cherche à caractériser la fonction $f^\star(.)$ à valeur continue dans $[0, 1]$ qui minimise l'espérance de la fonction coût $l(Y, f(X)) = |Y - f(X)|^q$ en fonction de q. Caractériser la solution pour $q = 2$ et $q = 1$ dans le cas discret et continu.

4. On pose L_{ij} le coût lorsqu'on assigne x à $\mathcal{C}_j$ alors que $x \in \mathcal{C}_i$.
— Exprimer la classe optimale $\mathcal{C}_i$ de x en fonction des $p(\mathcal{C}_j|x)$ et des L_{ij}.
— Dans le cas de deux classes, $\mathcal{C}_0$ et $\mathcal{C}_1$. On pose $p = p(\mathcal{C}_1|x)$, trouver le seuil de décision optimal $p^\star$ en fonction des L_{ij}.

5. Dans le tableau 2 figurent les résultats d'un classifieur multi-label sur une base de test.
— Quelles sont les fonctions de coût que l'on peut proposer pour quantifier les erreurs de prédictions en apprentissage multi-label ?
— Comment se géréralisent le rappel, la précision et la F-mesure au cas multi-label ? Faire l'application numérique à partir du taleau 2.

Y_1	1	0	1	1
Y_2	1	1	0	1
$\hat{Y}_1$	0	0	1	1
$\hat{Y}_2$	1	1	1	0

TABLE 2 – Prédictions obtenues sur la base de test

— Dans le tableau 3, figurent les probabilités conditionnelles $P(Y_1, Y_2|X)$ estimées par un classifieur probabiliste multilabel pour un X donné.

Y_1	Y_2	$\hat{P}(Y_1, Y_2\|X)$
0	0	0.25
0	1	0.30
1	0	0.27
1	1	0.18

TABLE 3 – Estimation de $P(Y_1, Y_2|X)$ par un classifieur probabiliste multilabel.

— Calculer la prédiction optimale avec le *Hamming loss*, le *zero-one loss*", le rappel, la précision et la F-mesure.

— Montrer que la solution qui minimise le *Hamming loss* ne dépend que des marginales $P(Y_j|X)$. En déduire une procédure simple d'apprentissage multilabel pour minimiser le *Hamming loss*.

6. Dans le tableau ci-dessous figurent les prédictions (ordonnées) d'un classfieur probabiliste avec les valeurs binaires cibles sur un jeu de test.

— Calculer le *rank loss*, la précision, le rappel, la F-mesure.

— Dessiner la courbe ROC associée à ces prédictions répétées. Que vaut l'AUC ?

— Si le coût de se tromper sur un exemple positif est 2 fois plus élévé que sur un exemple négatif, quels sont les seuils qui minimisent le coût moyen sur ce jeu de test ?

True Y	1	1	0	1	0	1	0	1	0	0
$\hat{P}(Y = 1 \mid X)$	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1

TABLE 4 – Prédictions d'un classfieur probabiliste sur une base de test.