

Initiation à la fouille de données - TD 1 - 09/2011

1. On sait qu'une pièce de monnaie est biaisée : $P(Pile) = 1/3$ et $P(Face) = 2/3$. A t-on intérêt à prédire qu'elle tombera toujours sur *Face* ou plutôt à prédire une fois sur 3 qu'elle tombera sur *Pile* et deux fois sur trois, qu'elle tombera sur *Face* ?
2. Un cas où la règle de décision de Bayes est calculable. Un sac contient une pièce équilibrée et deux pièces biaisées : $P(Pile) = 1/3$ et $P(Face) = 2/3$.
 - (a) On prélève une pièce, on la lance une fois. Quelle est la probabilité d'observer *Pile* ?
 - (b) On prélève une pièce, on la lance deux fois. Que doit-on décider si l'on observe *Pile* puis *Face* ? La pièce est-elle biaisée ou non ?
 - (c) On prélève une pièce, on la lance deux fois, on observe *Pile, Pile*. On la lance une troisième fois. Quelle est la probabilité d'observer *Pile* ?
 - (d) Une règle de décision doit associer une hypothèse (*biaisée, équilibrée*) à plusieurs observations : $\{PP, PF, FP, FF\}$. Calculez pour chaque observation la réponse donnée par la règle majoritaire, la règle du maximum de vraisemblance et la règle de Bayes. Quelle est la meilleure règle de décision ?
3. La règle de Bayes peut-elle définir le même classifieur que la règle majoritaire ?
4. Une entreprise de téléphonie a constaté que ses clients se répartissaient selon deux groupes : ceux qui téléphonent en moyenne aussi souvent vers un fixe que vers un portable et ceux qui téléphonent en moyenne deux fois plus souvent vers un portable que vers un fixe. Les clients du premier groupe sont environ deux fois plus nombreux que les seconds. L'entreprise souhaite adresser une offre promotionnelle aux clients du second groupe, pour éviter qu'ils ne changent de compagnie et ce, le plus tôt possible. Comme information sur sa clientèle, l'entreprise dispose des types d'appels effectués. On notera P (resp. F) un appel vers un portable (resp. vers un fixe).
 - (a) Peut-on décider sur la base des k premiers appels si un client appartient au premier groupe ou au second ? Supposons que $k = 4$ et qu'on observe la suite P, P, F, P . Quelle décision conduit à prendre la règle majoritaire ? la règle du maximum de vraisemblance ? la règle de Bayes ?
 - (b) Supposons que le bénéfice de l'opération soit mesuré selon la matrice suivante :

Prédit \ Réel	G_1	G_2
G_1	0	0
G_2	-1	2

Quelle décision l'observation de la suite P, P, F, P doit-elle entraîner ?