

IFT 603 : Devoir 4 - Questions additionnelles

Travail individuel

Remise : 10 avril 2015, 17h00 (**au plus tard**).

Remettez votre solution aux numéros 4 et 5 en format papier.

4. **[2.5 points]** Soit une variable aléatoire X pouvant prendre 3 valeurs : a , b ou c . Les probabilités associées à chacune de ces valeurs sont $p(X = a) = 0.2$, $p(X = b) = 0.7$ et $p(X = c) = 0.1$.

Supposez que vous ayez accès à un générateur de nombres aléatoires uniformément distribués entre 0 et 1, et qu'il ait généré la séquence suivante : 0.654, 0.234, 0.134, 0.993, 0.455. Convertissez ces 5 échantillons en 5 échantillons suivant la distribution de la variable aléatoire X définie ci-haut. Décrivez également la procédure que vous avez suivi.

5. **[2.5 points]** Une liste de décision (*decision list*) est un modèle de classification ayant la forme d'une série de déclarations "if ... else if ... else if ... else ...". Imaginons que les entrées \mathbf{x} d'un problème soient binaires (i.e. $x_i \in \{0, 1\} \forall i$) et considérons une liste de décision pour un classifieur binaire prenant la forme suivante :

```
if ( $x_a == A$ )  $\wedge$  ( $x_b == B$ ) then
  return +1
else if ( $x_c == C$ )  $\wedge$  ( $x_d == D$ ) then
  return +1
else if ( $x_e == E$ )  $\wedge$  ( $x_f == F$ ) then
  return +1
else
  return -1
end if
```

où $a, b, c, d, e, f \in \{1, \dots, D\}$, $a \neq b$, $c \neq d$, $e \neq f$ et $A, B, C, D, E, F \in \{0, 1\}$.

Dérivez une borne inférieure pour le nombre d'exemples m nécessaires afin d'assurer qu'avec probabilité d'au moins 0.95 un algorithme d'apprentissage cohérent (*consistent learner*) aura une erreur de généralisation d'au plus 0.1. Tentez de donner la meilleure borne possible et expliquez bien le raisonnement derrière cette borne.