

Analyse et Reconnaissance d'Images

James L. Crowley

M2R IVR

Premier Semestre 2007/2008

Séances 6 et 8

26 novembre et 10 décembre 2007

Exercice

Vous êtes analyste politique. Votre objectif est d'identifier les secteurs de la population pour une campagne ciblée de publicité en vue d'un référendum. Vous procédez par l'analyse des résultats d'un sondage. Chaque question de ce sondage comporte un ensemble fini de réponses. Les questions posées sont les suivantes :

- 1) Quel est votre âge ? : A) 18-29, B) 30-39 C) 40-49 D) 50-59, E) 60 ou plus
- 2) Quel est votre sexe ? : A) Masculin, B) Féminin
- 3) Quel est votre dernier Diplôme ? : A) Bac B) Bac+2, C) Bac+5 D) Doctorat E) Autre
- 4) Quel est votre salaire annuel ? : A) < 15 000 B) 15000 - 29999 C) 30 000 - 60 000 D) plus que 60 000.
- 5) Comment allez-vous voter ? A) "Oui" B) "Non C) Je ne suis pas décidé. D) Je ne voterai pas.

a) (2 points) Pour le groupe qui a répondu A ou B à la question 5, expliquez comment utiliser un ratio d'histogramme afin de prédire le vote le plus probable d'un électeur en fonction de son âge ? Combien de personnes faut-il avoir sondé ? Comment déterminer la probabilité d'erreur d'une telle prédiction ?

b) (2 points) Expliquez comment utiliser un ratio d'histogramme pour prédire la réponse à la question 5 à partir des réponses aux questions 1, 2, 3 et 4 ? Combien de personnes faut-il avoir sondé ? Comment déterminer la probabilité d'erreur ?

c) (2 points) Expliquez comment utiliser la fonction de discrimination quadratique vue en cours pour prédire la réponse à la question 5 à partir des réponses aux questions 1, 2, 3 et 4 ?

d) (2 points) Expliquez comment utiliser l'Algorithme EM pour découvrir les catégories des électeurs susceptibles de voter "Non" à partir des réponses aux questions 1, 2, 3 et 4 ?