Inférence Bayesienne du Maximum d'Entropie pour le Diagnostic du Cancer

F. Dornaika[*,**] et F. Chakik[***]

[*]IKERBASQUE, Basque Foundation for Science
[**]University of the Basque Country, San Sebastian, Spain
fadi_dornaika@ehu.es
[***]LaMA Laboratory, Lebanese University, Tripoli, Lebanon
fchakik@ul.edu.lb

1 Introduction et formulation

L'objectif de ce papier est de montrer que le principe du Maximum d'Entropie (Buck et Macaulay, 1991) emanant de la physique peut être utilisé en inférence statistique dans les tâches de classification binaires basées sur les exemples. Le Principe du Maximum d'Entropie est une approche systématique pour déterminer empiriquement la fonction de distribution de probabilités à partir de laquelle un ensemble de données a été tiré. Nous avons un ensemble d'apprentissage de M couples, $\{(\overrightarrow{x}_m, c_m)\}$, $m = 1, \ldots, M$, où $\overrightarrow{x}_m$ est un vecteur, que l'on appelle *exemple*, de dimension N, dont les composantes peuvent prendre des valeurs binaires ou réelles. Les c_m indiquent la classe de chaque exemple. Nous supposons qu'il y a C classes. Dans le cas d'une classification binaire, $C = 2$. La densité de probabilité qui maximise l'entropie a la forme générale suivante (Z est une constante de normalisation)

$$P(\overrightarrow{x}, c) = \frac{1}{Z} \exp\left[-\sum_n \lambda_n\, A_n(\overrightarrow{x}, c)\right] \tag{1}$$

où les $A_n(\overrightarrow{x}, c)$ sont des mesures "observables" qui sont fonction du vecteur $\overrightarrow{x}$ et de sa classe, les λ_n (scalaires ou vecteurs) sont à déterminer. Si l'on adopte les deux observables suivants : i) $\overrightarrow{A}_1(\overrightarrow{x}, c) = c\,\overrightarrow{x}$, et ii) $A_2(\overrightarrow{x}, c) = \overrightarrow{x}^2$, on obtiendra une solution analytique pour la densité : $P(\overrightarrow{x}, c) = \frac{1}{Z}\exp\left[-c\,\overrightarrow{x}\cdot\overrightarrow{\lambda}_1 - \lambda_2\,\overrightarrow{x}^2\right]$. Pour la classification binaire, on aura $c = \pm 1$. $\overrightarrow{\lambda}_1$ et λ_2 sont analytiquement estimés à partir des exemples (Chakik et al., 2004). Une fois la densité est connue, la classification adoptera la règle du Maximum a posteriori (MAP).

2 Diagnostic du cancer

Les exemples de cette application (Wolberg et Mangasarian, 1990) sont constitués par des vecteurs à 9 dimensions et qui sont classés comme bénins ou malins. Nous disposons, au