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 total, de 683 exemples dont 65.5% sont bénins. A partir de cet ensemble, nous générons deux ensembles d'apprentissage et de généralisation (test). Nous avons fixé la taille de l'ensemble d'apprentissage à 525 exemples, comme de nombreux auteurs, ce qui nous permettra de faire des comparaisons. Donc, l'ensemble de test est formé par les 158 exemples restants. Ainsi, nous avons estimé le taux de classification correcte en effectuant 10 partitions l'ensemble de données (10-fold cross validation). Les résultats sont tels que l'erreur d'apprentissage moyenne est de 3.37% alors que l'erreur de test moyenne est de l'ordre de 3.04%, et dont les variances sont de l'ordre de 0.07 et 2.1 respectivement. Ce résultat est légèrement moins bon que les résultats obtenus par *NetLines* (Torres et Gordon, 1998), dont l'erreur de test moyen est de l'ordre de 1.6%, mais meilleur que *Cascade Correlation* (Depenau, 1995), dont l'erreur de test est de 4.4%. La figure 1 illustre la sensibilité (qui dépend du taux de faux négatifs) et la spécificité (qui dépend du taux de faux positifs) correspondant aux 10 ensembles de test.

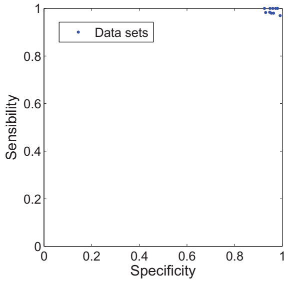

FIG. 1 – *Graphe représentant la sensibilité et la spécificité correspondant aux 10 ensembles de test.*

Références

Buck, B. et V. Macaulay (1991). *Maximum Entropy in Action*. Clarendon Press, Oxford.

Chakik, F., A. Shahin, J. Jaam, et A. Hasnah (2004). An approach for constructing complex discriminating surfaces based on bayesian interference of the maximum entropy. *International Journal of Information Sciences 163*(4).

Depenau, J. (1995). A global-local learning architecture for classification. In *International Neural Network Sociaety Meeting*.

Torres, J. et M. Gordon (1998). Efficient adaptive learning for classification tasks with binary units. *Neural Computation 10*(4), 1007–1030.

Wolberg, W. et O. Mangasarian (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. In *National Academy of Sciences*.

Summary

In this paper, we show that the Maximum Entropy Principle can be used by a Bayesian statistical inference for classification tasks. We present an application using benchmark data related to cancer diagnostic. Classification comparison is also provided.