Signaux Aléatoires

Classification Bayésienne

Jean-Yves Tourneret[1]

(1) University of Toulouse, ENSEEIHT-IRIT-TéSA
jyt@n7.fr

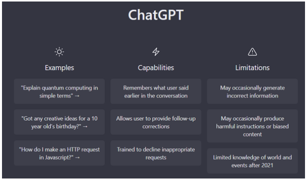
Année $2023 - 2024$

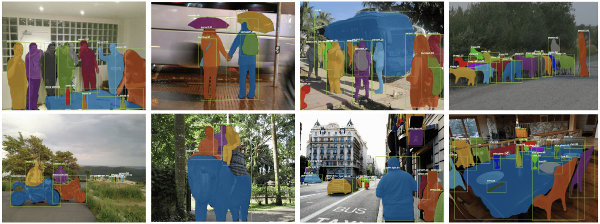
Bibliographie

Quelques références

- R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification, 2nd edition, Wiley, 2000.
- S. Theodoridis and K. Koutroumbas, Pattern Recognition, 4th edition, Academic Press, 2008.
- A. Jain, R. Duin and J. Mao, Statistical Pattern Recognition: A Review, IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 4-37, Jan. 2000.

Applications





Classification

Notations

- K classes $\omega_1, ..., \omega_K$
- $\boldsymbol{x} = [x(1), ..., x(p)]^T$ vecteur d'observations $\in X = \mathbb{R}^p$
- A: ensemble d'actions $a_1, ..., a_q$ où a_k "affecte le vecteur $\boldsymbol{x}$ dans la classe ω_k" avec $k = 1, ..., K$

Definition

$$d : \begin{array}{l} X \to A \\ \boldsymbol{x} \mapsto d(\boldsymbol{x}) \end{array}$$

Remarque

Classification avec reject: $A = \{a_0, a_1, ..., a_K\}$ où $a_0 =$ "on ne classe pas le vecteur $\boldsymbol{x}$"

Classification Bayésienne

Hypothèses : modèle probabiliste

- Probabilité *a priori* de la classe ω_k

$$P\left(\omega_k\right)$$

- Densité de probabilité du vecteur $\boldsymbol{x}$ conditionnelle à la classe ω_k

$$p\left(\boldsymbol{x}\left|\omega_k\right.\right)$$

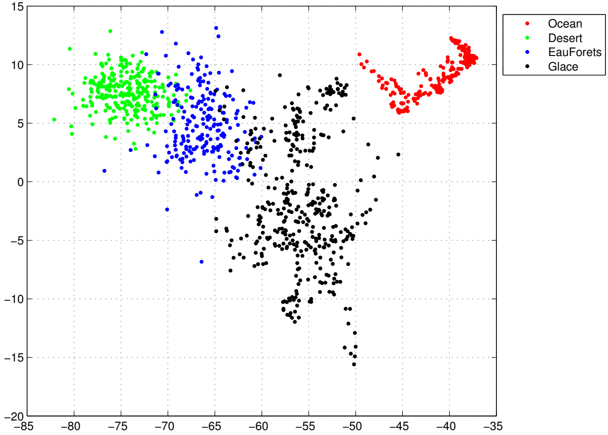
Conclusion

- Probabilité *a posteriori* que $\boldsymbol{x}$ appartienne à la classe ω_k

$$P\left(\omega_k\left|\boldsymbol{x}\right.\right) = \frac{p\left(\boldsymbol{x}\left|\omega_k\right.\right) P\left(\omega_k\right)}{p\left(\boldsymbol{x}\right)}$$

avec $p\left(\boldsymbol{x}\right) = \sum_{k=1}^{K} p\left(\boldsymbol{x}\left|\omega_k\right.\right) P\left(\omega_k\right)$.

Exemple



Ocean
Desert
EauForets
Glace
15
10
5
0
−5
−10
−15
−20
−85
−80
−75
−70
−65
−60
−55
−50
−45
−40
−35

Classifieur du maximum a posteriori (MAP)

Définition

$$d^*(\boldsymbol{x}) = a_j \Leftrightarrow P(\omega_j \,|\boldsymbol{x}) \geq P(\omega_k \,|\boldsymbol{x}), \forall k \in \{1, ..., K\}$$

Classes équiprobable : classifieur du maximum de vraisemblance

$$d^*(\boldsymbol{x}) = a_j \Leftrightarrow p(\boldsymbol{x} \,|\omega_j) \geq p(\boldsymbol{x} \,|\omega_k), \forall k \in \{1, ..., K\}$$

Propriété

Le classifieur MAP minimise la probabilité d'erreur de classification

Preuve (2 classes)

Probabilité d'erreur

$$\begin{aligned} P_e &= P\left[d\left(\boldsymbol{x}\right) = a_1 \cap \boldsymbol{x} \in \omega_2\right] + P\left[d\left(\boldsymbol{x}\right) = a_2 \cap \boldsymbol{x} \in \omega_1\right] \\ &= P\left[d\left(\boldsymbol{x}\right) = a_1 \mid \boldsymbol{x} \in \omega_2\right] P\left(\omega_2\right) + P\left[d\left(\boldsymbol{x}\right) = a_2 \mid \boldsymbol{x} \in \omega_1\right] P\left(\omega_1\right) \end{aligned}$$

Soit $R_i = \{\boldsymbol{x} \in \mathbb{R}^p / d(\boldsymbol{x}) = a_i\}$ la région associée à l'acceptation de la classe ω_i

$$\begin{aligned} P_e &= \int_{R_1} P\left(\omega_2\right) p\left(\boldsymbol{x}\left|\omega_2\right.\right) d\boldsymbol{x} + \int_{R_2} P\left(\omega_1\right) p\left(\boldsymbol{x}\left|\omega_1\right.\right) d\boldsymbol{x} \\ &= P\left(\omega_2\right)\left[1 - \int_{R_2} p\left(\boldsymbol{x}\left|\omega_2\right.\right) d\boldsymbol{x}\right] + \int_{R_2} P\left(\omega_1\right) p\left(\boldsymbol{x}\left|\omega_1\right.\right) d\boldsymbol{x} \\ &= P\left(\omega_2\right) + \int_{R_2} \left[P\left(\omega_1\right) p\left(\boldsymbol{x}\left|\omega_1\right.\right) - P\left(\omega_2\right) p\left(\boldsymbol{x}\left|\omega_2\right.\right)\right] d\boldsymbol{x} \\ &= P\left(\omega_2\right) - \int_{R_2} \left[P\left(\omega_2\left|\boldsymbol{x}\right.\right) - P\left(\omega_1\left|\boldsymbol{x}\right.\right)\right] p\left(\boldsymbol{x}\right) d\boldsymbol{x} \end{aligned}$$

P_e est minimum lorsque $R_2 = \{\boldsymbol{x} / P\left(\omega_2\left|\boldsymbol{x}\right.\right) > P\left(\omega_1\left|\boldsymbol{x}\right.\right)\}$

Probabilité d'erreur (K classes)

Définition

$$P_e = \sum_{k=1}^{K} P\left[d\left(\boldsymbol{x}\right) = a_k \cap \boldsymbol{x} \notin \omega_k\right]$$

Propriété

Le classifieur MAP minimise P_e

Preuve

admise

Cas gaussien

Densité

$$p\left(\boldsymbol{x}\,|\omega_k\right) = \frac{1}{(2\pi)^{p/2}\sqrt{\det \Sigma_k}} \exp\left[-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{m}_k\right)^T \boldsymbol{\Sigma}_k^{-1}\left(\boldsymbol{x}-\boldsymbol{m}_k\right)\right]$$

Cas général

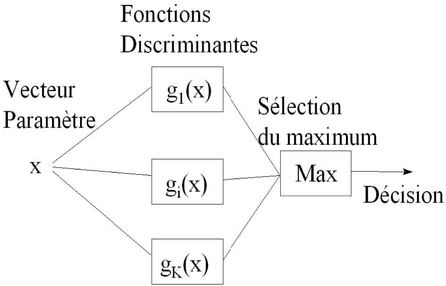
$$d^*\left(\boldsymbol{x}\right) = a_i \Leftrightarrow g_i\left(\boldsymbol{x}\right) \geq g_k\left(\boldsymbol{x}\right) \qquad \forall k = 1, ..., K$$

avec

$$g_i\left(\boldsymbol{x}\right) = -\left(\boldsymbol{x}-\boldsymbol{m}_i\right)^T \boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x}-\boldsymbol{m}_i\right) - \ln\det\boldsymbol{\Sigma}_i + 2\ln P\left(\omega_i\right)$$

Cas Gaussien

Classifieur



Fonctions discriminantes : Terme quadratique+ terme linéaire + constante

Matrices de covariance identiques ($\Sigma_i = \Sigma$)

Classes équiprobables : règles de la distance aux barycentres

$$d^*(\boldsymbol{x}) = a_i \Leftrightarrow d_M(\boldsymbol{x}, \boldsymbol{m}_i) \leq d_M(\boldsymbol{x}, \boldsymbol{m}_k) \qquad \forall k = 1, ..., K$$

où d_M est la distance de Mahalanobis

$$d_M(\boldsymbol{x}, \boldsymbol{m}_k) = \sqrt{(\boldsymbol{x} - \boldsymbol{m}_k)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{m}_k)}$$

Classes non équiprobables : fonctions discriminantes affines

$$d^*(\boldsymbol{x}) = a_i \Leftrightarrow \left[\boldsymbol{x} - \frac{1}{2}(\boldsymbol{m}_i + \boldsymbol{m}_k)\right]^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{m}_i - \boldsymbol{m}_k) \geq \ln \frac{P(\omega_k)}{P(\omega_i)}, \ \forall k$$

Que faut-il savoir ?

- Règle de classification Bayésienne
- Équivalence avec le classifieur du maximum de vraisemblance et de la règle de la distance aux barycentres
- Calcul de la probabilité d'erreur du classifieur Bayésien