UNIVERSITÉ DE MONTRÉAL

Rapport : Devoir1

Pierre Gérard
Mathieu Bouchard

IFT3395-6390 Fondements de l'apprentissage machine
Pascal Vincent, Alexandre de Brébisson et César Laurent

Année académique 2015 - 2016

Table des matières

1 Partie Théorique : Estimation de densité

1.1 Densité avec des fenêtres de Parzen a noyau Gaussien isotropique

a) $\mathcal{N}_{x^{(i)},\sigma^2} = \frac{1}{(2\pi)^{d/2}\sigma^d} e^{\frac{-1}{2}\frac{d(x^i,x)^2}{\sigma^2}}$

b) Si on choisit de prendre tous les points du voisinages $p(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{N}_{x^{(i)},\sigma^2}$ distribuer pour virer la somme ?

c) Hyper-paramètre : $\sigma \in \mathbb{R}$

Paramètre : On mémorise l'ensemble des données d'entrainement. ($n \times \mathbb{R}^d$)

On considère cette méthode comme non-paramétrique car le nombre de paramètre varie avec la taille de l'ensemble de données.

d) 3 propriété a prouvé : integral, integral=1, toujours positive (ou presque) Le petit d

1.2 Densité paramétrique avec Gaussienne isotropique

a) $p(x) = \frac{1}{(2\pi)^{d/2}\sigma^d} e^{\frac{-||x-\mu||^2}{2\Sigma^2}}$

μ est le est le vecteur des moyennes, de dimension d. Σ est la matrice de covariance, de dimension 1.

b) Dans le cas de l'estimation par fenêtre de Parzen, il nous faut effectuer une moyenne sur n densités Gaussiennes et, pour chacune d'elles, le point x(i) sert de moyenne $\mathcal{N}_{x^{(i)},\sigma^2}$ avec σ qui est un hyperparamètre. Dans le cas présent, on utilise une seule densité Gaussienne pour laquelle il faudra apprendre les valeurs de moyenne μ et d'écart-type σ.

c)

Paramètres :

- μ est le est le vecteur des moyennes, de dimension d.
- Σ est la matrice de covariance, de dimension 1.

Pas d'hyper-paramètre.

1.3 Densité paramétrique avec Gaussienne diagonale

a) $p(x) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} e^{\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

- μ est le est le vecteur des moyennes, de dimension d.
- Σ est la matrice de covariance, de dimension d.

b)Indépendance des composantes d'un vecteur aléatoire suivant une distribution Gaussienne diagonale :

Soit $x \in \mathbb{R}^d$ un vecteur aléatoire qui suit une distribution Gaussienne diagonale $\mathcal{N}_{\mu,\sigma(x)}$

Posons $x - \mu = k$ et $\frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} = s$

fin de la demo

c) gd c

d) gd d

1.4 Choix du modèle

La première tâche à effectuer pour choisir le modèle le plus approprié est de sélectionner un ensemble de tous les modèles que nous pourrions effectuer. Ici, il s'agit des approches présentées aux sections 1.1, 1.2 et 1.3. Ensuite, pour chacun des modèles sélectionnées, on procède ainsi :

1. Soit A , le modèle courant.
2. Si le modèle contient des hyper-paramètres, répéter la procédure pour toutes les valeurs acceptables que ces hyper-paramètres peuvent prendre. Appelons cette configuration λ. Rappelons que l'approche des fenêtres de Parzen à noyau isotropique contient un hyper- paramètre (l'écart-type σ) tandis que les approches 2 et 3 n'en n'ont aucun ($\lambda = \emptyset$).
3. Mélanger les éléments de l'ensemble D s'ils étaient ordonnés.

4. Séparer l'ensemble d'entrainement D en deux sous-ensembles D_{train} et D_{valid} . L'ensemble d'entrainement devrait être plus grand que l'ensemble de validation. $x \in \mathbb{R}$ qui suit une distribution Gaussienne

2 Partie Théorique : Classifieurs de Bayes

2.1 Classifieur de Bayes obtenu avec des densités paramétriques Gaussiennes diagonales

a)
b)
c)

2.2 Classifieur de Bayes obtenu avec des fenêtres de Parzen à noyau Gaussien isotropique

a)
b)

2.3 Classifieur de Bayes obtenu avec des fenêtres de Parzen à noyau Gaussien isotropique

a)
b)