Mean shift avec le noyau de Epanechnikov

Soit $P = \{p_1, \cdots, p_n\}$ un nuage de points dans l'espace euclidien $\mathbb{R}^d$. On note $\hat{f}_n \colon \mathbb{R}^d \to \mathbb{R}$ l'estimateur de densité à noyau de Epanechnikov défini à partir de P, de taille de fenêtre σ fixée.

Question 1 (Cours)**.** *Donnez la formule de $\hat{f}_n(x)$ en tout point $x \in \mathbb{R}^d$.*

Remarque. L'estimateur $\hat{f}_n$ n'est pas dérivable partout sur $\mathbb{R}^d$. Toutefois, ses points singuliers sont situés chacun à distance exactement σ d'au moins un point de P. Ainsi, l'ensemble des points singuliers de $\hat{f}_n$ est inclus dans l'union Σ des sphères centrées aux points de P et de même rayon σ, dont la mesure de Lebesgue est nulle. Dans toute la suite, pour simplifier les choses on travaillera sous l'hypothèse générique que $x \notin \Sigma$, ce qui garantira que $\hat{f}_n$ est dérivable en x.

Question 2. *Montrez que le gradient de $\hat{f}_n$ en tout point $x \notin \Sigma$ s'écrit comme suit :*

$$\nabla \hat{f}_n(x) = \frac{d+2}{\sigma^2}\, \hat{g}_n(x)\, (\bar{p}_\sigma(x) - x)\,,$$

où $\bar{p}_\sigma(x)$ désigne le centre de gravité des points $p_i \in P$ situés à distance au plus σ de x, et où $\hat{g}_n \colon \mathbb{R}^d \to \mathbb{R}$ désigne l'estimateur de densité à noyau uniforme de taille de fenêtre σ.

Ainsi, le gradient de l'estimateur de densité $\hat{f}_n$ en tout point $x \notin \Sigma$ est collinéaire au vecteur $\bar{p}_\sigma(x) - x$, connu sous le nom de *mean shift* de $\hat{f}_n$ en x. À la constante multiplicative $\frac{d+2}{\sigma^2}$ près, le coefficient scalaire appliqué au *mean shift* pour obtenir le gradient de $\hat{f}_n$ est donné par l'estimateur à noyau uniforme $\hat{g}_n$.

Question 3. *Montrez que pour tout point $x \notin \Sigma$ on a*

$$\hat{f}_n(\bar{p}_\sigma(x)) \geq \hat{f}_n(x),$$

avec égalité uniquement lorsque $\bar{p}_\sigma(x) = x$.

Ce résultat est à la base de la procédure dite de *mean shift hill-climbing* :

Input: $x \notin \Sigma$
  $x_0 := x$
  while $x_j \neq \bar{p}_\sigma(x_j)$ **do**
    $x_{j+1} := \bar{p}_\sigma(x_j)$
  end while

Remarque. En prolongement de la remarque précédente, dans toute la suite on fait l'hypothèse générique que les itérées x_j n'appartiennent pas à Σ et que, s'il existe, le point de convergence de la suite des itérées $(x_j)_{j\in\mathbb{N}}$ n'appartient pas non plus à Σ.

Question 4. *Déduisez de la question 3 que la suite des valeurs $\left(\hat{f}_n(x_j)\right)_{j\in\mathbb{N}}$ converge à la limite.*

Question 5. *Montrez que la convergence de la suite des valeurs $\left(\hat{f}_n(x_j)\right)_{j\in\mathbb{N}}$ se produit en réalité après un nombre fini d'itérations, et que la suite des itérées $(x_j)_{j\in\mathbb{N}}$ elle-même converge à ce moment-là.*

Question 6. *Montrez que, au point de convergence, le gradient de $\hat{f}_n$ est nul.*

Ainsi, sous notre hypothèse de généricité, la procédure de *mean shift hill-climbing* avec le noyau de Epanechnikov est garantie de converger en un nombre fini d'itérations vers un point critique de l'estimateur $\hat{f}_n$.

Mean shift avec le noyau de Epanechnikov

Question 1. On prend la formule générale d'un estimateur de densité à noyau radialement symétrique donnée dans le cours :

$$\hat{f}_n(x) \quad = \quad \frac{c_{k,d}}{n\,\sigma^d} \sum_{i=1}^{n} k\left(\frac{\|x-p_i\|_2^2}{\sigma^2}\right)$$

Ici k est le profil du noyau de Epanechnikov et la constant de normalisation $c_{k,d}$ est réglée en conséquence :

$$k(t) = \begin{cases} 1-t \text{ si } t \leq 1 \\ 0 \text{ si } t > 1 \end{cases} \quad = \max\{0,\, 1-t\} \qquad\qquad c_{k,d} = \frac{d+2}{2\,\mathrm{Vol}\,B_d(0,1)}$$

où $B_d(0,1)$ est la boule unité dans $\mathbb{R}^d$ et $\mathrm{Vol}\,B_d(0,1)$ est son volume. D'où :

$$\hat{f}_n(x) = \frac{d+2}{2n\,\sigma^d\,\mathrm{Vol}\,B_d(0,1)} \sum_{i=1}^{n} \max\left\{0,\, 1-\frac{\|x-p_i\|_2^2}{\sigma^2}\right\}$$

ce qui peut s'écrire alternativement comme suit :

$$\begin{aligned} \hat{f}_n(x) &= \frac{d+2}{2n\,\sigma^d\,\mathrm{Vol}\,B_d(0,1)} \sum_{i:\|x-p_i\|_2\leq\sigma} 1-\frac{\|x-p_i\|_2^2}{\sigma^2}, \text{ ou} \\ \hat{f}_n(x) &= \frac{d+2}{2n\,\sigma^d\,\mathrm{Vol}\,B_d(0,1)} \sum_{i:\|x-p_i\|_2<\sigma} 1-\frac{\|x-p_i\|_2^2}{\sigma^2} \end{aligned} \tag{1}$$

Ces trois formules sont valides comme résultat.

Remarque 1. Sous l'hypothèse générique de l'énoncé, on peut prendre n'importe laquelle des deux formules de l'équation (1) pour nos calculs ultérieurs car il n'y a aucun point de P à distance exactement σ de x. Dans la suite du corrigé on prend la première formule de l'équation (1).

Question 2. On dérive la première formule de l'équation (1) par rapport à x :

$$\begin{aligned} \nabla\hat{f}_n(x) &= -\frac{d+2}{2n\,\sigma^d\,\mathrm{Vol}\,B_d(0,1)} \sum_{i:\|x-p_i\|_2\leq\sigma} \frac{2}{\sigma^2}\,(x-p_i) \\ &= \frac{(d+2)\,\#\{i:\|x-p_i\|_2\leq\sigma\}}{n\,\sigma^{d+2}\,\mathrm{Vol}\,B_d(0,1)} \left(\frac{\sum_{i:\|x-p_i\|_2\leq\sigma} p_i}{\#\{i:\|x-p_i\|_2\leq\sigma\}} - x\right) \\ &= \frac{d+2}{\sigma^2} \left(\frac{1}{n\,\sigma^d\,\mathrm{Vol}\,B_d(0,1)} \sum_{i=1}^{n} \mathbb{1}_{\|x-p_i\|_2\leq\sigma}\right) (\bar{p}_\sigma(x)-x) \\ &= \frac{d+2}{\sigma^2} \left(\frac{1}{n\,\sigma^d\,\mathrm{Vol}\,B_d(0,1)} \sum_{i=1}^{n} \mathbb{1}_{\frac{\|x-p_i\|_2^2}{\sigma^2}\leq 1}\right) (\bar{p}_\sigma(x)-x) \\ &= \frac{d+2}{\sigma^2}\,\hat{g}_n(x)\,(\bar{p}_\sigma(x)-x) \end{aligned}$$

Question 3. On calcule :

$$\begin{aligned}
\frac{2n\,\sigma^d \operatorname{Vol} B_d(0,1)}{d+2}\,\hat{f}_n(\bar{p}_\sigma(x)) &= \sum_{i=1}^{n} \max\left\{0,\, 1 - \frac{\|\bar{p}_\sigma(x) - p_i\|_2^2}{\sigma^2}\right\} \\
&= \sum_{i:\|x-p_i\|_2\le\sigma}^{n} \max\left\{0,\, 1 - \frac{\|\bar{p}_\sigma(x) - p_i\|_2^2}{\sigma^2}\right\} \\
&\quad + \underbrace{\sum_{i:\|x-p_i\|_2>\sigma}^{n} \max\left\{0,\, 1 - \frac{\|\bar{p}_\sigma(x) - p_i\|_2^2}{\sigma^2}\right\}}_{\ge 0} \\
&\ge \sum_{i:\|x-p_i\|_2\le\sigma}^{n} \max\left\{0,\, 1 - \frac{\|\bar{p}_\sigma(x) - p_i\|_2^2}{\sigma^2}\right\} \\
&\ge \sum_{i:\|x-p_i\|_2\le\sigma}^{n} 1 - \frac{\|\bar{p}_\sigma(x) - p_i\|_2^2}{\sigma^2} \\
&= \#\left\{i : \|x-p_i\|_2 \le \sigma\right\} - \frac{1}{\sigma^2} \sum_{i:\|x-p_i\|_2\le\sigma}^{n} \|\bar{p}_\sigma(x) - p_i\|_2^2
\end{aligned}$$

À présent on peut utiliser un lemme vu à l'amphi sur k-means, qui dit que l'unique minimiseur de la variance dand $\mathbb{R}^d$ est la moyenne arithmétique. On l'applique à l'ensemble de points $\{p_i : \|x-p_i\|_2 \le \sigma\}$, ce qui donne l'inégalité :

$$\sum_{i:\|x-p_i\|_2\le\sigma}^{n} \|\bar{p}_\sigma(x) - p_i\|_2^2 \le \sum_{i:\|x-p_i\|_2\le\sigma}^{n} \|x - p_i\|_2^2$$

avec égalité uniquement lorsque $x = \bar{p}_\sigma(x)$. En conséquence, on a l'inégalité stricte suivante pour tout x tel que $x \neq \bar{p}_\sigma(x)$:

$$\begin{aligned}
\frac{2n\,\sigma^d \operatorname{Vol} B_d(0,1)}{d+2}\,\hat{f}_n(\bar{p}_\sigma(x)) &> \#\left\{i : \|x-p_i\|_2 \le \sigma\right\} - \frac{1}{\sigma^2} \sum_{i:\|x-p_i\|_2\le\sigma}^{n} \|x - p_i\|_2^2 \\
&= \sum_{i:\|x-p_i\|_2\le\sigma}^{n} \left(1 - \frac{\|x-p_i\|_2^2}{\sigma^2}\right) \\
&= \frac{2n\,\sigma^d \operatorname{Vol} B_d(0,1)}{d+2}\,\hat{f}_n(x)
\end{aligned}$$

Et pour tout x tel que $x = \bar{p}_\sigma(x)$ on a trivialement l'égalité $\hat{f}_n(\bar{p}_\sigma(x)) = \hat{f}_n(x)$.

Question 4. D'après la question 3 et l'hypothèse de généricité de l'énoncé, $\hat{f}_n(x_j)$ ne fait que croître durant la procédure. Par ailleurs, l'estimateur $\hat{f}_n$ est uniformément borné car chaque terme de la somme dans l'expression de $\hat{f}_n(x)$ est majoré par 1 :

$$\forall x \in \mathbb{R}^d, \quad \hat{f}_n(x) \le \frac{d+2}{2n\,\sigma^d \operatorname{Vol} B_d(0,1)} \sum_{i=1}^{n} 1 = \frac{d+2}{2\,\sigma^d \operatorname{Vol} B_d(0,1)}$$

qui est indépendant de x. En conséquence, la suite $\left(\hat{f}_n(x_j)\right)_{j\in\mathbb{N}}$ est croissante et majorée et donc converge.

Question 5. L'observation-clé est que, pour tout $x \notin \Sigma$, le point $\bar{p}_\sigma(x)$ se situe au centre de gravité d'un sous-ensemble de points de P. Comme P est fini, l'ensemble de ses parties l'est aussi, et donc l'ensemble $\{\bar{p}_\sigma(x) \mid x \notin \Sigma\}$ l'est également. La suite $\left(\hat{f}_n(x_j)\right)_{j\in\mathbb{N}}$ prend donc un nombre fini de valeurs distinctes. Comme elle est croissante par la question 3, elle converge au bout d'un nombre fini d'itérations.

Attention : cela ne signifie pas automatiquement que la procédure de *mean shift hill climbing* s'arrrête, puisque le critère d'arrêt est sur les itérées x_j, pas sur les valeurs $\hat{f}_n(x_j)$.

Après convergence de la suite $\left(\hat{f}_n(x_j)\right)_{j\in\mathbb{N}}$, on a $\hat{f}_n(x_{j+1}) = \hat{f}_n(x_j)$, c'est-à-dire $\hat{f}_n(\bar{p}_\sigma(x_j)) = \hat{f}_n(x_j)$, ce qui par la question 3 implique que $\bar{p}_\sigma(x_j) = x_j$, i.e. $x_{j+1} = x_j$. En d'autres termes, la suite des itérées $(x_j)_{j\in\mathbb{N}}$ elle-même converge et la procédure s'arrête.

Question 6. Par la question 2, au point de convergence x_j on a $\nabla\hat{f}_n(x_j)$ proportionnel au vecteur $(\bar{p}_\sigma(x_j) - x_j)$, qui est nul, par conséquent $\nabla\hat{f}_n(x_j) = 0$.