

TD 4

Méthodes à Noyaux

Ici log est le logarithme neperien.

Exercice 1

Pour chaque feature space \mathcal{X} et fonction $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, indiquer s'il s'agit d'un noyau symétrique positif (via une preuve) ou non (via un contre-exemple).

1) $\mathcal{X} = \{A, T, G, C\}^N$,

$$k : (x_1, \dots, x_N), (y_1, \dots, y_N) \mapsto \sum_{(t_1, t_2, t_3) \in \{A, T, G, C\}^3} \left(\sum_{k=0}^{N-3} \mathbb{1}(x_{k+1} = t_1, x_{k+2} = t_2, x_{k+3} = t_3) \right) \left(\sum_{k=0}^{N-3} \mathbb{1}(y_{k+1} = t_1, y_{k+2} = t_2, y_{k+3} = t_3) \right)$$

2) $\mathcal{X} = \mathbb{R}, k : x, y \mapsto \cos(x - y)$

3) $\mathcal{X} = \mathbb{R}, k : x, y \mapsto [0, +\infty)$ (une fonction symétrique positive)

4) $\mathcal{X} = [0, +\infty), k : x, y \mapsto \min(x, y)$

5) $\mathcal{X} = [0, +\infty), k : x, y \mapsto 1/\max(x, y)$

6) $\mathcal{X} = (-1, 1)^d, k : x, y \mapsto 1/(1 - x^T y)$

7) (bonus) $\mathcal{X} = \mathbb{N}_*, k : x, y \mapsto \text{pgcd}(x, y)$

Exercice 2

Soient $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{X} \times \{-1, 1\}$. Dans le cas où $\mathcal{X} = \mathbb{R}^d$, la régression logistique consiste à prédire $y = \text{sgn}(\hat{\theta}^T x)$ où sgn est la fonction de signe (prenant les valeurs ± 1) et où $\hat{\theta}$ est solution de

$$\hat{\theta} \in \underset{\theta \in \mathbb{R}^d}{\text{argmin}} \frac{1}{n} \sum_{i=1}^n \log(1 + e^{-y_i \theta^T x_i}) + \lambda \|\theta\|^2$$

où $\lambda \geq 0$ contrôle le niveau de régularisation. Le but de cet exercice est d'étudier la version "à noyaux" de cette méthode.

Étant donné \mathcal{H} un espace de Hilbert et $\phi : \mathcal{X} \mapsto \mathcal{H}$, nous allons considérer le noyau $k : x_1, x_2 \mapsto \langle x_1, x_2 \rangle_{\mathcal{H}}$.

1) Prouver que ce noyau est bien symétrique positif.

2) Justifier pourquoi nous pouvons nous restreindre à choisir un noyau symétrique positif sous cette forme.

La variante "à noyaux" de la régression logistique consiste à prédire $y = \text{sgn}(\langle \hat{\theta}, \phi(x) \rangle_{\mathcal{H}})$ où $\hat{\theta}$ est solution de

$$\hat{\theta} \in \underset{\theta \in \mathcal{H}}{\text{argmin}} \frac{1}{n} \sum_{i=1}^n \log(1 + e^{-y_i \langle \theta, \phi(x_i) \rangle_{\mathcal{H}}}) + \lambda \|\theta\|_{\mathcal{H}}^2$$

3) Supposons $\lambda > 0$, montrer que ce problème admet une unique solution $\hat{\theta}$ et que de plus $\hat{\theta} \in \text{Vect}(\phi(x_1), \dots, \phi(x_n))$.

Il est donc possible d'effectuer le changement de variables $\theta = \alpha_1 \phi(x_1) + \dots + \alpha_n \phi(x_n)$. La fonction de prédictions a alors la forme suivante :

$$f(x) = \text{sgn}(\langle \alpha_1 \phi(x_1), \phi(x) \rangle_{\mathcal{H}} + \dots + \langle \alpha_n \phi(x_n), \phi(x) \rangle_{\mathcal{H}}) .$$

4) Sur le jeu de données $\{(x_1, y_1) = ((0, 1), 1), (x_2, y_2) = ((-1, 1), -1), (x_3, y_3) = ((-1, -1), -1)\}$, représenter à peu près la frontière de décision lorsque $\alpha_1 = 1, \alpha_2 = \alpha_3 = -1$ pour un noyau Gaussien et pour le noyau linéaire.

- 5) Écrire le nouveau problème en $\alpha_1, \dots, \alpha_n$.
- 6) Sans se soucier de garantir la convergence, quel algorithme d'optimisation peut être utilisé pour résoudre le nouveau problème en $\alpha_1, \dots, \alpha_n$? Vous donnerez les formules des différentes quantités qui seront utiles pour le déroulement de l'algorithme.