Université d'Angers
Département de Mathématiques

Année 2018-2019
M2 Data Science

Statistique en Grande Dimension et Apprentissage - TD 4

Ce TD a pour but d'illustrer le chapitre 4 sur les SVM.

Exercice 1 (Hyperplans)**.**

1. Rappelez pourquoi l'équation d'un hyperplan affine $\mathcal{H}$ s'écrit
$$\langle \beta, x \rangle + \beta_0 = 0.$$
2. Montrez que pour un tel hyperplan $\mathcal{H}$ la distance d'un point x à $\mathcal{H}$ est égale à :
$$d(x, \mathcal{H}) = \frac{|\langle \beta, x \rangle + \beta_0|}{\|\beta\|}.$$
3. Quelle est l'équation d'un hyperplans parallèle à $\mathcal{H}$?

Exercice 2 (Optimisation/Lagrangien)**.** On se donne un échantillon $(x_i, y_i)_{i=1}^N$ où $y_i \in \{-1, 1\}$. On cherche à résoudre

$$\begin{cases} \min_{\beta,\beta_0} \frac{1}{2}\|\beta\|^2 \text{ sous la contrainte} \\ \forall i \in \{1, \dots, N\}, \quad y_i(\langle \beta, x_i \rangle + \beta_0) \geq 1. \end{cases} \tag{1}$$

1. Ecrire le Lagrangien $L(\beta, \beta_0, \alpha)$ associé à ce problème (où $\alpha = (\alpha_1, \dots, \alpha_N)$)
2. Ecrire les conditions de stationnarité : $\partial_\beta L = 0$ et $\partial_{\beta_0} L = 0$.
3. Notons $(\beta(\alpha), \beta_0(\alpha))$ un point critique de $(\beta, \beta_0) \mapsto L(\beta, \beta_0, \alpha)$. Pourquoi $(\beta(\alpha), \beta_0(\alpha))$ correspond-il à l'unique minimum de $(\beta, \beta_0) \mapsto L(\beta, \beta_0, \alpha)$?
4. Montrez que
$$\tilde{L}(\alpha) = L(\beta(\alpha), \beta_0(\alpha), \alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2}\sum_{i=1}^N \alpha_i y_i y_j \langle x_i, x_j \rangle.$$
On rappelle que le problème (1) se ramène au problème suivant :
$$\begin{cases} \max_{\alpha \in (\mathbb{R}^+)^N} \sum_{i=1}^N \alpha_i - \frac{1}{2}\sum_{i=1}^N \alpha_i y_i y_j \langle x_i, x_j \rangle \\ \text{avec } \sum_{i=1}^N \alpha_i y_i = 0. \end{cases}$$
Les conditions complémentaires de KKT garantissent de plus que si α^* désigne un tel maximiseur, alors pour tout i, soit $\alpha_i^* = 0$ soit $\alpha_i^* > 0$ et $y_i(\langle \beta^*, x_i \rangle + \beta_0^*) - 1 = 0$ (Les multiplicateurs de Lagrange non nuls correspondent aux vecteurs supports).
5. Montrez que $\beta^* = \sum_{i=1} \alpha_i^* y_i x_i$ puis que
$$\beta_0^* = -\frac{1}{2}\left(\min_{y_i=1}\langle \beta^*, x_i \rangle + \min_{y_i=-1}\langle \beta^*, x_i \rangle\right).$$

Exercice 3 (Noyau polynomial)**.** Soit $r \in \mathbb{N}$ tel que $r \geq 2$. On pose pour tous $x, x' \in \mathbb{R}^2$,

$$k(x, x') = (1 + \langle x, x' \rangle)^r.$$

1. Soit m un entier supérieur à 2. Montrez qu'il existe une application $\phi : \mathbb{R}^2 \to \mathbb{R}^{m+1}$ telle que pour tous $x, x' \in \mathbb{R}^2$,
$$\langle \phi(x), \phi(x') \rangle_{\mathbb{R}^{m+1}} = k(x, x').$$
2. Si on reprend la question précédente avec x, x' dans $\mathbb{R}^p$, quelle est la dimension de l'espace $\mathcal{H}$ associé (On pourra commencer par le cas $m = 2$) ?
3. Construire ϕ dans le cas où $k(x, x') = (1 + \langle x, x' \rangle)^r$ et $x, x' \in \mathbb{R}^2$.

Exercice 4 (Opérations sur les noyaux)**.** Montrez que la somme et le produit de deux noyaux sur un même espace est encore un noyau.

Exercice 5 (SVM Multi-Classes)**.** Différentes approches existent pour traiter les SVM multi-classes.

1. Approche globale : Ecrire un problème d'optimisation (primal) permettant de formuler les exigences d'un SVM linéaire (non flexible) dans le cas d'un problème à m classes.
2. Approche *un contre un* : on entraîne des SVM bi-classes puis on fabrique une règle globale. Sauriez-vous décrire plus en détail une telle approche ?
3. Approche *un contre tous* : A votre avis, à quoi cette approche correspond-elle ?

Exercice 6. Soient $\{(x_i, y_i), i \in \{1, \dots, N\}\}$ un échantillon inclus dans $\mathbb{R}^2 \times \{-1, 1\}$. On suppose que si $y_i = 1$, alors $x_i \in B(x_0, 1) = \{(a, b), (a - a_0)^2 + (b - b_0)^2 < 1\}$ $(x_0 = (a_0, b_0))$ et si $y_i > 1$, alors $x_i \in \{(a, b), (a - a_0)^2 + (b - b_0)^2 < 1\}$. En utilisant l'application ϕ définie par $\phi(a, b) = (a^2, b^2, a, b)$, montrez que les points sont linéairement séparables dans $\mathbb{R}^4$.