TD 10

Exercice 1 (6 points) – Couplage de SVMs

Soit deux problèmes de classification, le premier consistant à discriminer les images de chiffres entre des 1 et des 3, le deuxième entre des 3 et de 8. On dispose pour cela de deux ensembles d'apprentissage : $\mathcal{D}_1 = \{(\mathbf{x}_{i1}, y_{i1}), i1 = 1, \dots, n\}$ composé de 1 et de 3 et $\mathcal{D}_2 = \{(\mathbf{x}_{i2}, y_{i2}), i1 = 1, \dots, n\}$ composé de 3 et de 8. On souhaite apprendre deux classifieurs linéaires, le premier $f_1(\mathbf{x}) = \mathbf{w}_1 \cdot \mathbf{x}$ en charge de séparer les 1 des 3, le deuxième $f_2(\mathbf{x}) = \mathbf{w}_2 \cdot \mathbf{x}$ en charge de séparer les 3 des 8. Vu la similarité des problèmes, on sup pose que les solutions doivent être également similaires. Pour représenter cette similarité, on définit un paramètre $\mathbf{w}_0$ commun aux deux problèmes et les deux paramètres $\mathbf{w}_1$ et $\mathbf{w}_2$ comme des variations autour de ce paramètre :

$$\mathbf{w}_1 = \mathbf{w}_0 + \mathbf{v}_1 \text{ et } \mathbf{w}_2 = \mathbf{w}_0 + \mathbf{v}_2$$

En reprenant la formulation du SVM, la fonction à minimiser sous certaines contraintes est :

$$min_{\mathbf{w}_0,\mathbf{v}_1,\mathbf{v}_2,\xi_1,\xi_2} \frac{C_0}{2}(\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2) + \|\mathbf{w}_0\|^2 + C\sum_{i=1}^{n}(\xi_{i1} + \xi_{i2})$$

Q 1.1 A quoi servent les ξ_{ij} ? Quelles contraintes doit on poser pour compléter le problème de minimisation ?

Q 1.2 Que représente C et C_0 ? Comment les choisir ?

Q 1.3 Montrer que le Lagrangien peut s'écrire de la manière suivante et donner les contraintes sur $\alpha_{ij},\ \gamma_{ij}$

$$\mathcal{L} = \frac{1}{2}\|\mathbf{w}_0\|^2 + \frac{C_0}{2}\sum_{j=1}^{2}\|\mathbf{v}_j\|^2 + C\sum_{j=1}^{2}\sum_{i=1}^{n}\xi_{ij} - \sum_{j=1}^{2}\sum_{i=1}^{n}\alpha_{ij}[y_{ij}(\mathbf{w}_0 + \mathbf{v}_j)\cdot\mathbf{x}_{ij} - 1 + \xi_{ij}] - \sum_{j=1}^{2}\sum_{i=1}^{n}\gamma_{ij}\xi_{ij}$$

Q 1.4 Quelles sont les conditions d'optimalité par rapport aux variables primales $\mathbf{w}_0$, $\mathbf{v}_j$ et ξ_{ij} ?

Q 1.5 Quelles sont les expressions de $\mathbf{w}_0$ et $\mathbf{v}_j$?

Q 1.6 Donner un encadrement des α_{ij}.

Q 1.7 Quel est le problème dual correspondant, en particulier l'expression de $\mathcal{L}$ sans $\mathbf{w}_0,\ \mathbf{v}_j$?

Q 1.8 Exprimer $f_1(\mathbf{x})$ et $f_2(\mathbf{x})$ en fonction des α_{ij}.

Exercice 2 (5 points) – Déphasé

On considère le modèle suivant $y = f(x) + \epsilon$ avec $f(x) = \sum_{j=1}^{d} w_j sin(cx + \phi_j)$, avec $\mathbf{w} \in \mathbb{R}^d$, $\boldsymbol{\phi} \in \mathbb{R}^d$, $c, x \in \mathbb{R}$ et on considère que ϵ suit une loi normale centrée en 0 et de variance $1/\alpha^2$.

Q 2.1 Donner un argument pour montrer que y suit une loi normale quand on fixe x, c, $\mathbf{w}$, $\boldsymbol{\phi}$ et α. Donner l'expression de $p(y|x, c, \alpha, \mathbf{w}, \boldsymbol{\phi})$.

Q 2.2 Soit un ensemble de données $\mathcal{D} = \{x^i, y^i\}_{i=1}^{n}$, quelle est la vraisemblance du modèle par rapport à $\mathcal{D}$? Donner également l'expression de la log-vraisemblance.

Q 2.3 Donner un algorithme pour optimiser $\mathbf{w}$ lorsque tous les autres paramètres sont connus.

Q 2.4 (bonus) Et si on voulait également apprendre les $\boldsymbol{\phi}$, quelle approche préconiseriez vous ? Discuter des problèmes potentiels.