Assignment 3

Due 2014/11/05 23:59:59

Submit your work with the filename: "`ML_assignment-${number}_${your-student-id}_${your-name}`." Late submissions will not be accepted.

d = data dimension   N = data size

1. Shown that in regularized linear regression, $\boldsymbol{w}^* = (\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I}_d)^{-1} \boldsymbol{X}^\top \boldsymbol{r} = \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{X}^\top + \lambda \boldsymbol{I}_N)^{-1} \boldsymbol{r}$, a linear combination of the examples.
(Hint: use the matrix inversion in block form: `http://www.cs.nthu.edu.tw/~jang/book/addenda/matinv/matinv/`)

2. Show that in an RKHS, the inner product $\langle f, g\rangle := \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha^{(i)} \beta^{(j)} k(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(j)})$ for any $f = \sum_{i=1}^{n} \alpha^{(i)} k(\boldsymbol{x}^{(i)}, \cdot)$ and $g = \sum_{j=1}^{m} \beta^{(j)} k(\boldsymbol{y}^{(j)}, \cdot)$ is well-defined; i.e., it satisfies

   (a) *symmetry*: $\langle f, g\rangle = \langle g, f\rangle$;

   (b) *linearity*: $\langle af + bg, h\rangle = a\langle f, h\rangle + b\langle g, h\rangle$ for any $a, b \in \mathbb{R}$; and

   (c) *positive definiteness*: $\langle f, f\rangle \geq 0$ with equality holds iff $f(\cdot) = 0(\cdot)$.

3. Show that in a large-margin linear classifier, the margin between the hyperplanes $\{\boldsymbol{x} : \boldsymbol{w}^\top \boldsymbol{x} - b = 1\}$ and $\{\boldsymbol{x} : \boldsymbol{w}^\top \boldsymbol{x} - b = -1\}$ is $2/\|\boldsymbol{w}\|$.

4. Prove the Semiparametric Representer theorem.

5. Consider the necessary optimality condition for the SVM dual (that is, if the following holds, then $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$):

$$\exists \rho \in \mathbb{R} \text{ such that } \max_{i \in I_{up}} r^{(i)} g_i \leq \rho \leq \min_{j \in I_{down}} r^{(j)} g_j \tag{1}$$

   where $I_{up} := \{i | r^{(i)} \alpha_i < B^{(i)}\}$ and $I_{down} := \{j | A^{(j)} < r^{(j)} \alpha_j\}$.

   (a) Show that the above condition can be rewritten as

$$\exists \rho \in \mathbb{R} \text{ such that } \forall t, \begin{cases} r^{(t)} \alpha_t^* = B^{(t)}, & \text{if } g_t^* > r^{(t)} \rho \\ r^{(t)} \alpha_t^* = A^{(t)}, & \text{if } g_t^* < r^{(t)} \rho \end{cases}, \tag{2}$$

   which is equivalent to

$$\exists \rho \in \mathbb{R} \text{ such that } \forall t, \begin{cases} \alpha_t^* = C, & \text{if } g_t^* > r^{(t)} \rho \\ \alpha_t^* = 0, & \text{if } g_t^* < r^{(t)} \rho \end{cases}. \tag{3}$$

(b) Show that Condition (1) is also sufficient by letting $\boldsymbol{w}^* = \sum_{t=1}^{N} \alpha_t^* r^{(t)} \Phi(\boldsymbol{x}^{(t)})$, $b^* = \rho$, and $\xi_t^* = \max\{0, g_t^* - r^{(t)}\rho\}$.
(Hint: show that there is no duality gap, i.e., $primal(\boldsymbol{w}^*, b^*, \boldsymbol{\xi}^*) - dual(\boldsymbol{\alpha}^*) = 0$, by Condition (3))

6. Show that the *Area Under the ROC Curve* (AUC) is equal to the probability that a classifier ranks a randomly chosen positive instance higher than a randomly chosen negative one.
(Hint: by partitioning the AUC horizontally)