模式识别 6 陈斯杰、 2016310721

Problem 4 Programming

训练样本为 $(X, Y) = \{(x_1, y_1), \dots, (x_n, y_n)\}$, $x_i \in R^D$, $y_i \in R$

线性预测器: $\hat{y}(X, w) = W^T \phi(x)$ 其中 $\phi(x) = (\phi_1(x), \phi_2(x), \dots, \phi_M(x))^T$ 为M个特征.

$$J(w, \lambda) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}(y_i - W^T\phi(x))^2 + \lambda \|W\|_1$$

4.1 Show that the sub-differential of $J(w,\lambda)$ with respect to w_k is:

$$\partial_{w_k} J(w,\lambda) = (a_k w_k - c_k) + \lambda \partial_{w_k} |w_k|$$

$$= \begin{cases} \{a_k w_k - (c_k + \lambda)\}, & w_k < 0 \\ [-c_k - \lambda, \ -c_k + \lambda], & w_k = 0 \\ \{a_k w_k - (c_k - \lambda)\}, & w_k > 0 \end{cases}$$

其中,

$$a_k = \frac{1}{n}\sum_{i=1}^{n} \phi_k^2(x_i)$$

$$c_k = \frac{1}{n}\sum_{i=1}^{n} \phi_k(x_i)(y_i - W_{-k}^T \phi_{-k}(x_i))$$

求 $J(w,\lambda)$ 对 w_k 的 sub-differential 可分两步:

$$\partial_{w_k} J(w,\lambda) = \partial_{w_k}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(y_i - w^T\phi(x))^2\right] + \partial_{w_k}[\lambda \|W\|_1]$$

$$\partial_{w_k}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(y_i - w^T\phi(x))^2\right]$$

$$\cancel{\frac{1}{n}\sum_{i=1}^{n}[y_i - w^T\phi(x)]\cdot\partial_{w_k}[y_i - w^T\phi(x)]}$$

$$= \partial_{w_k}\frac{1}{2n}\sum_{i=1}^{n}[y_i - W_{-k}^T\phi_{-k}(x_i) - w_k\phi_k(x_i)]^2$$

$$= \frac{1}{2n}\sum_{i=1}^{n}\partial_{w_k}\left[(y_i - W_{-k}^T\phi_{-k}(x_i))^2 + w_k^2\phi_k^2(x_i) - 2w_k\phi_k(x_i)(y_i - W_{-k}^T\phi_{-k}(x_i))\right]$$

$$= \frac{1}{2n}\left[\sum_{i=1}^{n} 2w_k\phi_k^2(x_i) - 2\phi_k(x_i)(y_i - W_{-k}^T\phi_{-k}(x_i))\right]$$

$$= w_k \cdot \frac{1}{n}\sum_{i=1}^{n}\phi_k^2(x_i) - \frac{1}{n}\sum_{i=1}^{n}[\phi_k(x_i)(y_i - W_{-k}^T\phi_{-k}(x_i))]$$

$$= a_k w_k - c_k$$

$$\partial_{w_k}[\lambda\|W\|_1] = \lambda \cdot \partial_{w_k}\sum_{i=1}^{M}|w_i|$$

$$= \lambda\, \partial_{w_k}|w_k|$$

$$= \begin{cases} \{-\lambda\} & w_k < 0 \\ [\lambda, -\lambda] & w_k = 0 \\ \{\lambda\} & w_k > 0 \end{cases}$$

模式识别6 陈斯杰，2016310721

综上：

$$\partial_{w_k} J(w,\lambda) = \begin{cases} \{a_k w_k - \lambda - c_k\} & w_k < 0 \\ [-c_k - \lambda,\ -c_k + \lambda] & w_k = 0 \\ \{a_k w_k + \lambda - c_k\} & w_k > 0 \end{cases}$$

4.2: Solve the optimality condition

(a) $c_k < -\lambda$ 时，

$\lambda + c_k > 0$，$a_k w_k - \lambda - c_k$ 表示的直线截距大于0，

$a_k w_k - (c_k + \lambda) = 0 \Rightarrow w_k = \dfrac{c_k + \lambda}{a_k} < 0$

(b) $c_k \in [-\lambda, \lambda]$ 时，

$w_k = 0$

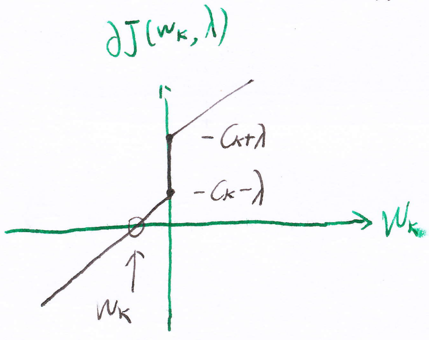
(c) $c_k > \lambda$ 时，

$\lambda - c_k < 0$，$a_k w_k + \lambda - c_k$ 表示的截距小于0

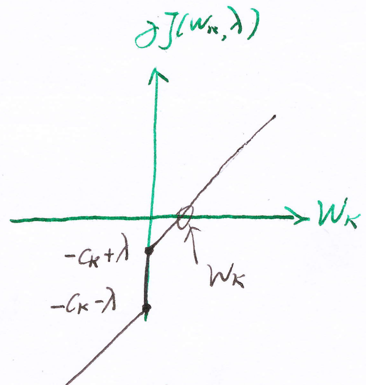
$a_k w_k - (c_k - \lambda) = 0 \Rightarrow w_k = \dfrac{c_k - \lambda}{a_k} > 0$

综上：

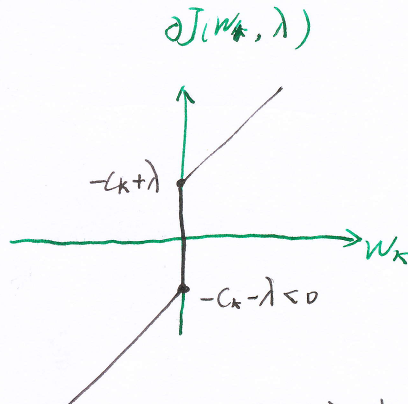
$$w_k = \begin{cases} \dfrac{c_k + \lambda}{a_k} & c_k < -\lambda \\ 0 & c_k \in [-\lambda, \lambda] \\ \dfrac{c_k - \lambda}{a_k} & c_k > \lambda \end{cases}$$



$c_k < -\lambda$ 时



$c_k > \lambda$ 时



$c_k \in [-\lambda, \lambda]$ 时

w_k 时 λ 的关系