4. Consider a regression problem, whereby, we are given feature vectors $\{\mathbf{x}_i \in \mathbb{R}^d\}$ and response variables $\{y_i \in \mathbb{R}\}$. The objective is to minimize the error between the estimated and true response variables. In order to control overfitting, we add a regularization term. The problem can be formulated as follows:

$$\begin{aligned} \underset{\mathbf{w},\boldsymbol{\xi}}{\text{minimize}} \quad & L = \sum_{i=1}^{n} \xi_i^2 \\ \text{subject to} \quad & y_i - \mathbf{w}^T\mathbf{x} = \xi_i,\ \forall i = 1,2,\dots n \\ & \|w\|_2 \le B. \end{aligned}$$

Here, B is the regularization parameter.

(a) Obtain a solution of the problem by rewriting it in dual form.

(b) Does this problem have the equivalent of support vectors as in SVMs? Justify.

(c) What is one basic disadvantage of the above as compared to the SVM solution?

[30+25+5=60]

$$\mathcal{L}(w, \xi, u, \lambda) = \sum \xi_i^2 + \sum u_i(y_i - w^T x_i - \xi_i) + \lambda(w^T w - B^2) \quad - ①$$

$\frac{\partial \mathcal{L}}{\partial w} = 0$ $\qquad \sum u_i x_i = 2\lambda w$ $\qquad \boxed{w = \frac{1}{2\lambda}\sum u_i x_i}$

$\frac{\partial \mathcal{L}}{\partial \xi_i} = 0$ $\qquad 2\xi_i = u_i$ $\qquad \xi_i = \frac{u_i}{2}$

On Replacing in ①

$$\mathcal{L}(u, \lambda) = \sum \frac{u_i^2}{4} + \sum u_i y_i - \sum \frac{u_i^2}{2} - \frac{\sum\sum u_i u_j \langle x_i, x_j\rangle}{2\lambda} + \lambda \frac{\sum\sum u_i u_j \langle x_i, x_j\rangle}{4\lambda^2} - \lambda B^2$$

$$= -\frac{u_i^2}{4} + \sum_i u_i y_i - \frac{\sum_i\sum_j u_i u_j \langle x_i, x_j\rangle}{4\lambda} - \lambda B^2$$

$$\boxed{\mathcal{L}(u,\lambda) = -\frac{u^T u}{4} + u^T y - \frac{u^T X^T X u}{4\lambda} - \lambda B^2}$$ Dual Problem

Dual Problem

$$\max_{u,\lambda} \mathcal{L}(u,\lambda) = -\frac{u^T u}{4} + u^T y - \frac{u^T X^T X u}{4\lambda} - \lambda B^2 \qquad \lambda \ge 0, \quad \lambda(w^T w - B^2) = 0$$

$$\Rightarrow \min_{u,\lambda} \mathcal{L}(u,\lambda) = \frac{u^T u}{4} - u^T y + \frac{u^T X^T X u}{4\lambda} + \lambda B^2$$

$\frac{\partial \mathcal{L}}{\partial u} = 0$ $\qquad \frac{u}{2} - y + \frac{X^T X u}{2\lambda} = 0$ $\qquad \boxed{u = 2\left[I + \frac{X^T X}{\lambda}\right]^{-1} y}$

$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$ $\qquad \frac{u^T X^T X u}{4\lambda^2} = B^2 \quad \Rightarrow \quad \boxed{u^T X^T X u = 4\lambda^2 B^2}$

$\overline{\partial u}$

$$\frac{\partial L}{\partial \lambda} = 0 \qquad \frac{u^T X^T X u}{4\lambda^2} = B^2 \quad \Rightarrow \quad \boxed{u^T X^T X u = 4\lambda^2 B^2}$$

2λ

$$w = \frac{\sum u_i x_i}{2\lambda} = \frac{X u}{2\lambda} = \frac{X}{\cancel{2\lambda}} \frac{\cancel{2}}{\cancel{\lambda}} (\lambda I + X^T X)^{-1} \cancel{y}$$

$$w = X (\lambda I + X^T X)^{-1} y$$

(b) $w = \frac{\sum u_i x_i}{2\lambda}$ Like SVM w is a linear sum of x_i,

Thus it has an equivalent of support vectors

However in this case all would be support vectors, as $u_i \neq 0$

(c) Unlike the case of SVM's where only a small fraction of vectors formed the set of support vectors. Here each vector is a SV themselves. This method is thus not robust to outliers and noisy data

Also the dual problem is a convex optimisation problem in 2 variables, more difficult to be accurately solved.