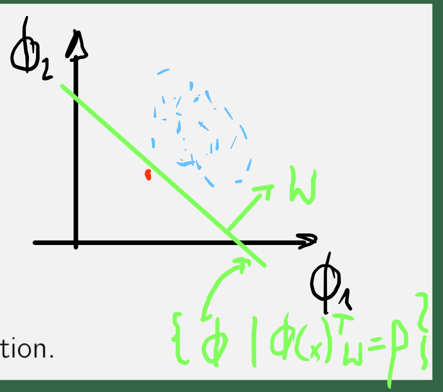
Exercise 1: One-Class SVM ($5+5+20+10+10$ P)

The one-class SVM is given by the minimization problem:



$$\min_{\boldsymbol{w},\rho,\xi} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 - \rho + \frac{1}{N\nu}\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \quad \forall_{i=1}^{N} : \langle \phi(\boldsymbol{x}_i), \boldsymbol{w}\rangle \geq \rho - \xi_i \quad \text{and} \quad \xi_i \geq 0$$

where $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are the training data and $\phi(\boldsymbol{x}_i) \in \mathbb{R}^d$ is a feature space representation.

(a) *Show* that strong duality holds (i.e. verify the Slater's conditions).

we can always increase ξ_i until the constraints are satisfied with strict inequalities.

(b) *Write* the Lagrange function associated to this optimization problem.

$$\mathcal{L}(w,\rho,\xi,\alpha,\beta) = \frac{1}{2}\|w\|^2 - \rho + \frac{1}{N\nu}\sum_i \xi_i + \sum_i \alpha_i\left(\rho - \xi_i - \langle \phi(x_i), w\rangle\right) - \sum_i \beta_i \xi_i$$

(c) *Show* the dual program for the one-class SVM is given by:

$$\max_{\boldsymbol{\alpha}} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j k(\boldsymbol{x}_i, \boldsymbol{x}_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i = 1 \quad \text{and} \quad \forall_{i=1}^{N} : \ 0 \leq \alpha_i \leq \frac{1}{N\nu}$$

$$\max_{\alpha,\beta} \ \min_{w,\rho,\xi} \ \mathcal{L}(w,\rho,\xi,\alpha,\beta)$$

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_i \alpha_i \phi(x_i) \overset{!}{=} 0 \quad \Rightarrow \quad w = \sum_i \alpha_i \phi(x_i)$$

$$\frac{\partial \mathcal{L}}{\partial \rho} = \sum_i \alpha_i - 1 \overset{!}{=} 0 \quad \Rightarrow \quad \sum_i \alpha_i = 1$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = \frac{1}{N\nu} - \alpha_i - \beta_i \overset{!}{=} 0 \quad \text{and} \quad \beta_i \geq 0 \quad \Rightarrow \frac{1}{N\nu} - \alpha_i \geq 0$$

$$\Rightarrow \frac{1}{N\nu} \geq \alpha_i$$

$$\mathcal{L}(w,\rho,\xi,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \cancel{\sum_i \xi_i\left(\frac{1}{N\nu} - \alpha_i - \beta_i\right)} + \cancel{\rho\left(\sum_i \alpha_i - 1\right)} - \underbrace{\langle \sum_i \alpha_i \phi(x_i), w\rangle}_{\|w\|^2}$$

$$= \frac{1}{2}\|w\|^2 - \|w\|^2 = -\frac{1}{2}\|w\|^2$$

$$= -\frac{1}{2}\left\|\sum_i \alpha_i \phi(x_i)\right\|^2 = -\frac{1}{2}\langle \sum_i \alpha_i \phi(x_i), \sum_j \alpha_j \phi(x_j)\rangle$$

$$= -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j k(x_i, x_j) \quad \rightsquigarrow \quad \max_{\alpha} \mathcal{L}(\alpha) \quad \text{s.t.} \quad 0 \leq \alpha_i \leq \frac{1}{N\nu}, \ \sum_i \alpha_i = 1$$

(d) *Show* that the problem can be equivalently rewritten in canonical matrix form as:

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\boldsymbol{\alpha}^\top K \boldsymbol{\alpha}$$

$$\text{s.t.} \quad 1^\top \boldsymbol{\alpha} = 1 \quad \text{and} \quad \begin{pmatrix} -I \\ I \end{pmatrix} \boldsymbol{\alpha} \preceq \begin{pmatrix} 0 \\ 1/N\nu \end{pmatrix}$$

where K is the Gram matrix whose elements are defined as $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

$$\max_{\alpha} \ -\alpha^T K \alpha \quad \rightsquigarrow \quad \min_{\alpha} \ \alpha^T K \alpha$$

$$1^T \alpha = \sum_i 1_i \alpha_i = \sum_i \alpha_i = 1$$

$$\begin{bmatrix} -\alpha_1 \\ \vdots \\ -\alpha_N \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \preceq \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1/N\nu \\ \vdots \\ 1/N\nu \end{bmatrix} \quad \rightsquigarrow \quad \begin{matrix} -\alpha_i \leq 0 \\ \alpha_i \leq 1/N\nu \end{matrix}$$

(e) The decision rule in the primal for classifying a point as an outlier is given by:

$$\langle \phi(\boldsymbol{x}), \boldsymbol{w}\rangle < \rho$$

Also, one can verify that for any data point $\boldsymbol{x}_i$ whose associated dual variable satisfies the strict inequalities $0 < \alpha_i < \frac{1}{N\nu}$, and calling one such point a support vector $\boldsymbol{x}_{\text{SV}}$, the following equality holds:

$$\langle \phi(\boldsymbol{x}_{\text{SV}}), \boldsymbol{w}\rangle = \rho$$

Show that the outlier detection rule can be expressed as:

$$\sum_{i=1}^{N}\alpha_i k(\boldsymbol{x}, \boldsymbol{x}_i) < \sum_{i=1}^{N}\alpha_i k(\boldsymbol{x}_{\text{SV}}, \boldsymbol{x}_i)$$

$$\langle \phi(x), w\rangle < \rho$$

$$\Leftrightarrow \langle \phi(x), w\rangle < \langle \phi(x_{SV}), w\rangle$$

$$\Leftrightarrow \langle \phi(x), \sum_i \alpha_i \phi(x_i)\rangle < \langle \phi(x_{SV}), \sum_i \alpha_i \phi(x_i)\rangle$$

$$\Leftrightarrow \sum_i \alpha_i k(x, x_i) < \sum_i \alpha_i k(x_{SV}, x_i)$$