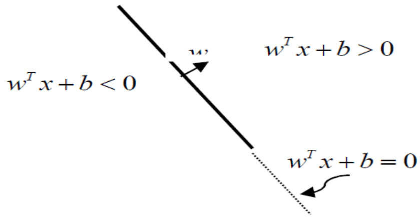
Tutorial 10 : SVM

1 Theory

Plane Geometry - Remainder

- The projection of a vector x at the direction of a vector w is $\frac{w^T x}{||w||_2}$.
- Equation of a plane in $\mathbb{R}^d$ is $w^T x$ where $b \in \mathbb{R}$ and $w \in \mathbb{R}^d$ are constant which define the plane. It can be seen that cb and cW for some constant $c \in \mathbb{R}$ define the same plane.
- Euclidean distance of a point x_0 from a plane defined by b and w is $d(x_0) = \frac{w^T x_0 + b}{||w||_2}$. The sign of d determines whether the point is on the side of the plane which is parallel to w or on the side which is anti-parallel to w:



Support Vector Machine (SVM)

Primal Problem (P)

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||_2^2 \\ s.t. \quad & y_k(w^T x_k + b) \geq 1, \; k = 1, 2, ..., n. \end{aligned}$$

Dual Problem (D)

$$\begin{aligned} \max_{\alpha} \quad & \sum_{k=1}^{n} \alpha_k - \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} \alpha_k \alpha_l y_k y_l \langle x_k, x_l \rangle \\ s.t. \quad & \alpha_k \geq 0, \; k = 1, 2, ..., n, \\ & \sum_{k=1}^{n} \alpha_k y_k = 0. \end{aligned}$$

Support Vectors

1. Their distance to the separating plane is minimal.
2. $\alpha_k > 0$ if and only if x_k is a support vector.
3. If x_k is a support vector then $y_k(w^T x_k + b) = 1$ (The other direction is not necessarily true since there might be an example x_j for which the equality above holds but $\alpha_j = 0$.
4. The euclidean distance of a support vector to the separating plane is called the *margin* of the problem and it is equal to $\frac{1}{||w||_2}$.

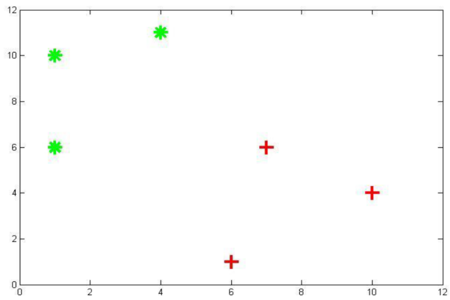
Practice

Question 1

The following two classes are given

Class 1: $[1, 6]$, $[1, 10]$, $[4, 11]$ for $y_k = -1$.

Class 2: $[6, 1]$, $[7, 6]$, $[10, 4]$ for $y_k = +1$.



(a) What are probably the support vectors?

(b) It is given that the optimal values of the dual problem are

$$\alpha = [0.0356\ 0\ 0.04\ 0.0756\ 0]^T.$$

To which samples belong the zero values?

(c) Find the optimal value of the vector w of the primal problem. What is the margin of the problem?

Solution 1

(a) This is a simple problem for which it is easy to see that the support vectors are $[1, 6]$, $[4, 11]$ and $[7, 6]$.

(b) The zero values belong to the samples which are not support vectors given by $[1, 10]$, $[6, 1]$ and $[10, 4]$.

(c) Recall from the lecture that $w = \sum_{k=1}^{n} \alpha_k y_k x_k$, hence, we get that

$$w = 0.0356 \cdot (-1) \cdot [1, 6] + 0.04 \cdot (-1) \cdot [4, 11] + 0.0756 \cdot (+1) \cdot [7, 6] = \frac{1}{15}[5, -3].$$

The margin of the problem is given by

$$\frac{1}{||w||_2} = 2.5725.$$

Question 2

Define $\xi_k = \max\Big(0, 1 - y_k(w^T x_k + b)\Big)$, we can rewrite the soft-SVM problem as the following optimization problem:

$$\begin{aligned} \min_{w\in\mathbb{R}^d, \xi\in\mathbb{R}^n, b\in\mathbb{R}} \ & \frac{1}{2}||w||_2^2 + C\sum_{k=1}^{n} \xi_k \\ s.t.\quad & y_k(w^T x_k + b) \geq 1 - \xi_k, \qquad k = 1, 2, ..., n, \\ & \xi_k \geq 0, \qquad\qquad\qquad\quad k = 1, 2, ..., n. \end{aligned} \tag{1}$$

Denote the optimal solution to problem (1) by (w^*, b^*, ξ^*).
Given $\delta > 0$, we define a new problem as

$$\begin{aligned} \min_{w\in\mathbb{R}^d,\xi\in\mathbb{R}^n,b\in\mathbb{R}} & \frac{1}{2}||w||_2^2 + \delta C\sum_{k=1}^{n}\xi_k \\ s.t. \quad & y_k(w^T x_k + b) \geq \delta - \xi_k, \quad k = 1, 2, ..., n, \\ & \xi_k \geq 0, \quad k = 1, 2, ..., n. \end{aligned} \tag{2}$$

Define $(\hat{w}, \hat{b}, \hat{\xi}) \triangleq (\delta w^*, \delta b^*, \delta \xi^*)$.

(a) Prove that $(\hat{w}, \hat{b}, \hat{\xi})$ is a feasible solution to problem (2), i.e., it satisfies the inequality constraints.

(b) Prove that $(\hat{w}, \hat{b}, \hat{\xi})$ is the optimal solution to problem (2).

(c) Prove that the optimal solution to (2) leads to the same classification as the optimal solution to (1), i.e.,

$$sign\big(\langle w^*, x\rangle + b^*\big) = sign\big(\langle \hat{w}, x\rangle + \hat{b}\big).$$

Solution

(a) The optimal solution to (1) satisfies

$$\begin{aligned} & y_k\big(\langle w^*, x\rangle + b^*\big) \geq 1 - \xi_k^*, \quad k = 1, 2, ..., n, \\ & \xi_k^* \geq 0, \quad k = 1, 2, ..., n. \end{aligned}$$

Multiplying both inequalities by $\delta > 0$ we get

$$\begin{aligned} & y_k\big(\langle \delta w^*, x\rangle + \delta b^*\big) \geq \delta - \delta\xi_k^*, \quad k = 1, 2, ..., n, \\ & \delta\xi_k^* \geq 0, \quad k = 1, 2, ..., n. \\ & \Downarrow \\ & y_k\big(\langle \hat{w}, x\rangle + \hat{b}\big) \geq \delta - \hat{\xi}_k, \quad k = 1, 2, ..., n, \\ & \hat{\xi}_k \geq 0, \quad k = 1, 2, ..., n. \end{aligned}$$

Therefore, $(\hat{w}, \hat{b}, \hat{\xi})$ is a feasible solution to problem (2).

(b) Assume by contradiction that there exists a solution $\big(\tilde{w}, \tilde{b}, \tilde{\xi}\big)$ which achieves a value of the objective function which is strictly smaller than the value achieved by $(\hat{w}, \hat{b}, \hat{\xi})$.

Claim: $(\frac{\tilde{w}}{\delta}, \frac{\tilde{b}}{\delta}, \frac{\tilde{\xi}}{\delta})$ is a feasible solution to problem (1).

__**Proof**__

$\big(\tilde{w}, \tilde{b}, \tilde{\xi}\big)$ is a feasible solution to problem (2), therefore,

$$\begin{aligned} & y_k\big(\langle \tilde{w}, x\rangle + \tilde{b}\big) \geq \delta - \tilde{\xi}_k, \quad k = 1, 2, ..., n, \\ & \tilde{\xi}_k \geq 0, \quad k = 1, 2, ..., n. \\ & \Downarrow \\ & y_k\big(\langle \frac{\tilde{w}}{\delta}, x\rangle + \frac{\tilde{b}}{\delta}\big) \geq 1 - \frac{\hat{\xi}_k}{\delta}, \quad k = 1, 2, ..., n, \\ & \frac{\tilde{\xi}_k}{\delta} \geq 0, \quad k = 1, 2, ..., n. \end{aligned}$$

By our assumption, we have that

$$\begin{aligned}
&\frac{1}{2}||\tilde{w}||_2^2 + \delta C \sum_{k=1}^{n} \tilde{\xi}_k < \frac{1}{2}||\hat{w}||_2^2 + \delta C \sum_{k=1}^{n} \hat{\xi}_k \\
\Rightarrow &\frac{1}{2}||\tilde{w}||_2^2 + \delta C \sum_{k=1}^{n} \tilde{\xi}_k < \frac{1}{2}||\delta w^*||_2^2 + \delta C \sum_{k=1}^{n} \delta \xi_k^* \\
\Rightarrow &\frac{1}{2}||\tilde{w}||_2^2 + \delta C \sum_{k=1}^{n} \tilde{\xi}_k < \delta^2 \left(\frac{1}{2}||w^*||_2^2 + C \sum_{k=1}^{n} \xi_k^* \right) \\
&\frac{1}{2}\left|\left|\frac{\tilde{w}}{\delta}\right|\right|_2^2 + C \sum_{k=1}^{n} \frac{\tilde{\xi}_k}{\delta} < \frac{1}{2}||w^*||_2^2 + C \sum_{k=1}^{n} \xi_k^*.
\end{aligned}$$

Therefore, $(\frac{\tilde{w}}{\delta}, \frac{\tilde{b}}{\delta}, \frac{\tilde{\xi}}{\delta})$ achieves a lower value of the objective function which is in contradiction to the optimality of (w^*, b^*, ξ^*).

(c)

$$\begin{aligned}
sign\big(\langle \hat{w}, x \rangle + \hat{b}\big) &= sign\big(\langle \delta w^*, x \rangle + \delta b^*\big) \\
&= sign(\delta) sign\big(\langle w^*, x \rangle + b^*\big) \\
&= sign\big(\langle w^*, x \rangle + b^*\big).
\end{aligned}$$