Example

$$\min_{x \in \mathbb{R}} f_0(x) = x^2 \quad \text{subject to} \quad f_1(x) = ax + b \leq 0$$

$$L(x, y) = f_0(x) + y_1 f_1(x) \quad \text{for } x \in \mathbb{R},\ y \geq 0$$

$$\tilde{L}(y) = \inf_{x \in \mathbb{R}} f_0(x) + y_1 f_1(x)$$

$$\frac{d}{dx}\left(f_0(x) + y_1 f_1(x)\right) = 0$$

$$2x + y_1 a = 0 \Leftrightarrow x = -y_1 \frac{a}{2}$$

$$\left(f_0(x) + y_1 f_1(x)\right)\Big|_{x = -y_1 \frac{a}{2}}$$

$$= y_1^2 \frac{a^2}{4} - y_1^2 \frac{a}{2} + y_1 b$$

$$= -y_1^2 \frac{a^2}{4} + y_1 b = \tilde{L}(y)$$

We do not need to check $|x| \to \infty$

$$\sup_{y_1 \geq 0} \tilde{L}(y) = ?$$

$$\frac{d}{dy_1}\left(-y_1^2 \frac{a^2}{4} + y_1 b\right) = -y_1 \frac{a^2}{2} + b = 0$$

$$\Leftrightarrow \quad y_1 = \frac{2b}{a^2}$$

$$\sup_{y \geq 0} \tilde{L}(y) = \begin{cases} \frac{2b}{a^2} & \text{for } b \geq 0 \Rightarrow x = -\frac{b}{a} \\ 0 & \text{for } b < 0 \Rightarrow x = 0 \end{cases}$$

PRIMAL & DUAL FORM OF SVM

I

Primal program

$$f_0(w,\xi) = \frac{1}{2}\|\underline{w}\|^2 + C\sum_{i=1}^{m} \xi_{i} \longrightarrow \min$$

subject to constraints

$$f_i(w,\xi) = 1 - y^{(i)}\, w\cdot x^{(i)} - \xi_{(i)} \le 0 \qquad i=1\dots m$$

$$g_j(w,\xi) = -\xi_{(j)} \le 0 \qquad j=1\dots m$$

$\Rightarrow$ Lagrange function

$$L(w,\xi,\alpha,\beta) = f_0(w,\xi) + \sum_{i=1}^{m} \alpha_i f_i(w,\xi) + \sum_{j=1}^{m} \beta_j g_j(w,\xi)$$

at optimum w,ξ,α,β we have KKT:

(I) $$\begin{pmatrix}\partial_{w_0}\\ \partial_{\underline{w}} \\ \partial_{\xi_k}\end{pmatrix} \left[f_0(w,\xi) + \sum_{i=1}^{m} \alpha_i f_i(w,\xi) + \sum_{j=1}^{m} \beta_j g_j(w,\xi)\right] = 0$$

(II) $\alpha_i f_i(w,\xi) = 0$, $\beta_j g_j(w,\xi) = 0$, $\alpha_i \ge 0$, $\beta_i \ge 0$

for $i,j = 1\dots m$

$\partial_{w_0} f_0(w,\xi) = 0$, $\partial_{\underline{w}} f_0(w,\xi) = \underline{w}$, $\partial_{\xi_k} f_0(w,\xi) = C$

$\partial_{w_0} f_i(w,\xi) = -y^{(i)}$, $\partial_{\underline{w}} f_i(w,\xi) = -y^{(i)}\underline{x}^{(i)}$, $\partial_{\xi_k} f_i(w,\xi) = -\delta_{ik}$

$\partial_{w_0} g_j(w,\xi) = 0$, $\partial_{\underline{w}} g_j(w,\xi) = 0$, $\partial_{\xi_k} g_j(w,\xi) = -\delta_{jk}$

II

(I) $\Longleftrightarrow$ $$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

$$\underline{w} = \sum_{i=1}^{m} \alpha_i y^{(i)} \underline{x}^{(i)}$$

$$C = \alpha_i + \beta_i \qquad i=1\dots m$$

(II) $\Longleftrightarrow$ $$\alpha_i\left(1 - y^{(i)}\, w\cdot x^{(i)} - \xi_{(i)}\right) = 0 \qquad \alpha_i \ge 0 \qquad i=1\dots m$$

$$\beta_j \xi_j = 0 \qquad \beta_j \ge 0 \qquad j=1\dots m$$

We note again as in the lin. sep. case that due to (I)

$$\underline{w} = \sum_{i=1}^{m} \alpha_i y^{(i)} \underline{x}^{(i)}$$

where due to (II) only vectors fulfilling $\alpha_i \neq 0$, i.e.,

$$1 - y^{(i)}\, w\cdot x^{(i)} - \xi^{(i)} = 0$$

contribute $\Rightarrow$ either support vectors with $\xi^{(i)} = 0$ or other vectors for $\xi^{(i)} \neq 0$. The latter however must satisfy $\beta_i \xi^{(i)} = 0 \Rightarrow \beta_i = 0 \Rightarrow \alpha_i = C$.

Hence, vectors contributing to $\underline{w}$ are either support vectors or outliers with $\alpha_i = C$.

Note however that even though w is unique as we have shown, the support vectors are not.

Dual program: Thanks to the KKT Theorem

we know that $\exists\ \bar{\alpha}, \bar{\beta}$ s.t.

$$\inf \{ f_0(w, \xi) \mid f_i(w, \xi) \le 0,\ g_j(w, \xi) \le 0,\ i = 1 \ldots n,\ j = 1 \ldots n \}$$

$$= \inf_{w, \xi} L(w, \xi, \bar{\alpha}, \bar{\beta}) = \max_{\alpha, \beta \ge 0} \inf_{w, \xi} L(w, \xi, \alpha, \beta)$$

Furthermore, by KKT cond. we know how to express

$$\underline{w}^* = \sum_{i=1}^{n} \alpha_i y^{(i)} \underline{x}^{(i)}$$

$$\Rightarrow L(w^*, \xi, \alpha, \beta) = \frac{1}{2} \left\| \sum_{i=1}^{n} \alpha_i y^{(i)} \underline{x}^{(i)} \right\|^2 + C \sum_{i=1}^{n} \xi_i$$
$$+ \sum_{i=1}^{n} \alpha_i \left(1 - y^{(i)} \left[w_0 + \sum_{j=1}^{n} \alpha_j y^{(j)} \underline{x}^{(j)} \right] - \xi_i \right)$$
$$+ \sum_{i=1}^{n} \beta_i (-\xi_i)$$
$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \underline{x}^{(i)} \cdot \underline{x}^{(j)}$$
$$+ \sum_{i=1}^{n} [C - \alpha_i - \beta_i] \xi_i - w_0 \sum_{i=1}^{n} \alpha_i y^{(i)}$$

Hence, we may formulate an equivalent problem

$$\tilde{f}_0(\alpha, \beta) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \underline{x}^{(i)} \cdot \underline{x}^{(j)} \longrightarrow \max$$

subject to $\alpha_i + \beta_i = C,\ \alpha_i, \beta_i \ge 0 \iff 0 \le \alpha_i \le C$

and $\sum_i \alpha_i y^{(i)} = 0$

NOTE: This program has several advantages:

1) $\partial_\alpha^2 \tilde{f}_0(\alpha) = - y^{(i)} \underline{x}^{(i)} \cdot y^{(j)} \underline{x}^{(j)}$ which is negative semi-def.

$\Rightarrow \tilde{f}_0(\alpha)$ concave

2) Constraints are all affine

$\Rightarrow \exists!$ optimum and problem is always quadratic

We can furthermore express the activation output

$$h(x) = wx$$

by choosing one support vector $\bar{x}^{(*)}$, i.e., $w \bar{x}^{(*)} = \bar{y}^{(*)}$ and computing

$$w_0 = \bar{y}^{(*)} - \underline{w} \cdot \underline{\bar{x}}^{(*)}$$
$$= \bar{y}^{(*)} - \sum_{j=1}^{n} \alpha_j y^{(j)} \underline{x}^{(j)} \underline{\bar{x}}^{(*)}$$

$\Rightarrow$ The activation does not depend on the particular values of $x^{(i)}$ but only on inner products.

- **Activation function can be readily computed in dual form.**