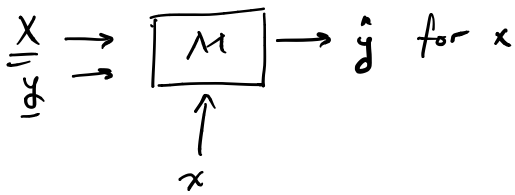
Outline: Duality

Kernel SVM



$$\hat{w} = (X^T X)^{-1} X^T y \Rightarrow \hat{y} = \hat{w}^T x$$

Constrained optimization

$$\min_x \quad f(x)$$

$$\text{s.t.} \quad g_i(x) \le 0 \quad i = 1, \dots, m$$

$$h_j(x) = 0 \quad j = 1, \dots, n$$

Lagrangian: $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{n} \mu_j h_j(x)$

$\underline{\lambda} = [\lambda_1, \dots, \lambda_m]$, $\underline{\mu} = [\mu_1, \dots, \mu_n]$ are dual variables

Lagrange Dual function

$$L_D(\lambda, \mu) = \min_x \; L(x, \lambda, \mu)$$

$L_D(\lambda, \mu)$ is concave (pointwise min. affine functions)

$$\min_x \left\{ \max_{\substack{\lambda_i \ge 0 \;\forall i \\ \mu_j \in \mathbb{R} \;\forall j}} L(x, \lambda, \mu) \right\} = \min_x \begin{cases} f(x) & g_i(x) \le 0,\ h_j(x) = 0 \ \forall i, j \\ \infty & \text{if } x \text{ is not feasible} \end{cases}$$

$$= \min_{\substack{x:\ g_i(x) \le 0\ \forall i \\ h_j(x) = 0\ \forall j}} f(x)$$

Dual optimization problem

$$\max_{\substack{\lambda \ge 0 \\ \mu \in \mathbb{R}^n}} \min_x L(x, \lambda, \mu) = \max_{\substack{\lambda \ge 0 \\ \mu \in \mathbb{R}}} L_D(\lambda, \mu)$$

Weak Duality:

$$p^* = \min_x \max_{\substack{\lambda \ge 0 \\ \mu}} L(x, \lambda, \mu)$$

Theorem: $d^* \le p^*$.

$$d^* = \max_{\substack{\lambda \ge 0 \\ \mu}} \min_x L(x, \lambda, \mu)$$

Proof: If x' feasible $\quad g_i(x') \le 0 \quad h_j(x') = 0 \quad \forall i,j$

$$L(x', \lambda, \mu) = f(x') + \sum \overbrace{\lambda_i g_i(x')}^{\le 0} + \sum \mu_j \cancel{h_j(x')}^{0} \le f(x')$$

$$\underbrace{\min_x L(x, \lambda, \mu)}_{L_D} \overset{(\text{min.})}{\le} L(x', \lambda, \mu) \le f(x') \quad \text{for any } x' \text{ feasible}$$

$$L_D \le \min_{\substack{x':\ g_i(x') \le 0\ \forall i \\ h_j(x) = 0\ \forall j}} f(x') = p^*$$

$$\Rightarrow d^* = \max_{\substack{\lambda \ge 0 \\ \mu}} L_D(\lambda, \mu) \le p^* \qquad \square$$

Strong Duality: If $p^* = d^*$, we say strong duality holds

Theorem: If $f, g_1, \ldots, g_m$ are convex, $h_1, \ldots, h_n$ are affine, and a constraint qualification holds, then $p^* = d^*$.

Examples of constraint qualification

Slater's Condition: $\exists x$ s.t. $g_i(x) < 0$ and $h_j(x) = 0$ $\forall i, j$.

E.g. $0 \le x_i \le 1$ $x \in \mathbb{R}^d$ Slater's cond holds

$$\frac{x_1^2}{x_2} \le 0 \qquad x_1 \in \mathbb{R}, \; x_2 \in \mathbb{R}_+$$

Slater's Condition does not hold! (implies $x_1 = 0$)

Exercise:
$$\min \; e^{-x_1}$$
$$\frac{x_1^2}{x_2} \le 0$$
$$x_2 \ge 0$$

Show that Strong duality does not hold.

KKT Conditions

Assume $f, g_1, \ldots, g_m$ are diff'ble $h_1, \ldots, h_n$ affine

Unconstrained case $\nabla f(x^*) = 0$

Thm: If $p^* = d^*$, x^* is primal optimal, (λ^*, μ^*) dual optimal, then Karush-Kuhn-Tucker conditions hold

1) Stationarity $\nabla_x L(x, \lambda, \mu)$ at (x^*, λ^*, μ^*)

2) Primal Feasible $g_i(x^*) \le 0$ $\forall i$ $h_j(x) = 0$ $\forall j$

3) Dual Feasible $\lambda_i^* \ge 0$ $\forall i$

4) Complimentary Slackness $\lambda_i^* g_i(x^*) = 0 \quad \forall i$

$$\left(\text{if } g_i(x^*) < 0 \Rightarrow \lambda_i^* = 0 \right)$$

Theorem: If $f, g_1, \dots, g_m$ convex, $h_1, \dots, h_n$ affine and (x^*, λ^*, μ^*) satisfy KKT, then (x^*, λ^*, μ^*) is optimal and strong duality holds.

Exercise: $\min\limits_{w} \frac{1}{2} \| Xw - y \|_2^2 + \frac{1}{2} \lambda \cdot \|w\|_2^2 \quad \left(w^* = (X^T X + \lambda I)^{-1} \cdot X^T y \right)$

$$= \min_{\substack{z, w \\ z = Xw}} \frac{1}{2} \| z - y \|_2^2 + \frac{1}{2} \lambda \cdot \|w\|_2^2$$

$$= \min_{w, z} \max_{\alpha} \alpha^T (z - Xw) + \frac{1}{2} \|z - y\|_2^2 + \frac{\lambda}{2} \cdot \|w\|_2^2$$

Strong duality $\rightarrow$ $= \max\limits_{\alpha} \min\limits_{w, z} \alpha^T (z - Xw) + \frac{1}{2} \|z - y\|_2^2 + \frac{\lambda}{2} \cdot \|w\|_2^2$

$$-X^T \alpha + \lambda \cdot w = 0 \Rightarrow \boxed{w^* = \frac{X^T \alpha}{\lambda}}$$

$$\alpha + z - y = 0 \Rightarrow z^* = y - \alpha$$

$$= \max_{\alpha} \alpha^T \cdot (y - \alpha) - \frac{\alpha^T \cdot X \cdot X^T \alpha}{\lambda} + \frac{1}{2} \cdot \|\alpha\|_2^2 + \frac{\lambda}{2} \cdot \left\| \frac{X^T \alpha}{\lambda} \right\|_2^2$$

$$= \max_{\alpha} \alpha^T y - \frac{1}{2} \|\alpha\|_2^2 - \frac{1}{2\lambda} \cdot \alpha^T X X^T \alpha$$

$$\nabla_\alpha (\ ) = 0 \Rightarrow y - \alpha - \frac{1}{\lambda} \cdot X X^T \alpha = 0$$

$$\Rightarrow \alpha^* = \left(\frac{1}{\lambda} X X^T + I \right)^{-1} y.$$

$$w^* = \frac{X^T \alpha^*}{\lambda} = \frac{X^T}{\lambda} \left(\frac{1}{\lambda} X X^T + I \right)^{-1} y$$

$K = XX^T$
Kernel matrix

Given x : $\quad w^{*T} x = \left(\frac{1}{\lambda} \sum_{i=1}^{n} x_i \cdot \alpha_i^* \right)^T x$

$$= \frac{1}{\lambda} \sum_{i=1}^{n} \underbrace{\langle x_i, x \rangle}_{k'(x)} \cdot \alpha_i^*$$

Kernel Ridge Reg. $\quad \alpha^* = \left(\frac{1}{\lambda} K + I \right)^{-1} y$

and prediction $\quad \frac{1}{\lambda} \sum_{i=1}^{n} k(x_i, x) \cdot \alpha_i^*.$

Soft-margin SVM

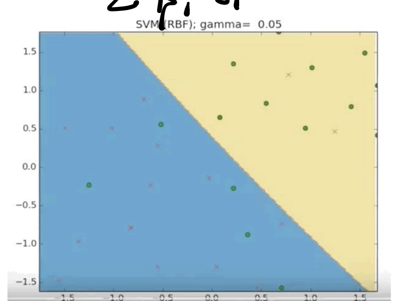
$$\min_{w, s_i, b} \ \frac{1}{2} \|w\|_2^2 + \frac{C}{n} \sum s_i$$

$$y_i (w^T x_i + b) \geq 1 - s_i \quad \forall i$$

$$s_i \geq 0 \quad \forall i$$

$$L(w, b, s_i, \alpha, \beta) = \frac{1}{2} \|w\|_2^2 + \frac{C}{n} \sum_{i=1}^{n} s_i - \sum \alpha_i (y_i (w^T x_i + b) + s_i - 1)$$

$$- \sum \beta_i \cdot s_i$$



Dual problem: $\quad \max_{\substack{\alpha \geq 0 \\ \beta \geq 0}} L_D(\alpha, \beta)$

$$L_D(\alpha, \beta) = \min_{w, b, s_i} L(w, b, s_i, \alpha, \beta)$$

$$\frac{\partial L}{\partial w} = 0 = w - \sum_{i=1}^{n} \alpha_i y_i x_i \Rightarrow \boxed{w^* = \sum_{i=1}^{n} \alpha_i \cdot y_i \underline{x_i}}$$

$$\frac{\partial L}{\partial b} = 0 = \sum^{n} \alpha_i y_i = 0$$

$\hat{y} = \text{sign}(w^{*T} x + b^*)$

$\frac{\partial L}{\partial b} = 0 = \sum_{i=1} \alpha_i y_i$

$= \text{sign}\left(\sum \alpha_i^* y_i \langle x_i, x\rangle + b^*\right)$

$= \text{sign}\left(\sum \alpha_i^* y_i \, k(x_i, x) + b^*\right)$

$e^{-\frac{\|x_i - x\|_2^2}{2\sigma^2}}$

$\frac{\partial L}{\partial s_i} = 0 = \frac{C}{n} - \alpha_i - \beta_i = 0$

$$L_D(\alpha, \beta) = -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum \alpha_i$$

Dual problem $\max_{\alpha, \beta} \; -\frac{1}{2} \sum \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum \alpha_i$

s.t. $\sum \alpha_i y_i = 0$

$\alpha_i + \beta_i = \frac{C}{n}$

$\alpha_i \geq 0, \; \beta_i \geq 0$