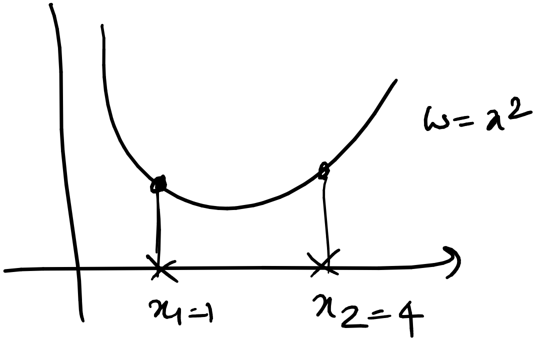
Tool:

1. Convex functions $f(\underline{\theta x + (1-\theta) y}) \le \theta f(x) + (1-\theta) f(y)$

$\theta \in [0,1]$

$\theta = 40\%$
$= .4$



2. $\min\limits_{x} f(x) = f^* \quad (x^*)$

subject $h_i(x) \le 0 \quad i = 1, \dots m$

$l_j(x) = 0 \quad j = 1, \dots r$

(Primal)

Key idea: (Primal) problem (Dual) problem

Lagrangian.

$$\mathcal{L}(x, \underbrace{u}_{\mathbb{R}^m_+}, \underbrace{v}_{\mathbb{R}^r}) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x)$$

dual variables

$$C = \left\{ x : \begin{array}{l} h_i(x) \le 0 \\ l_j(x) = 0 \end{array} \right\}$$

$$f^\star \geq \min_{x \in C} \mathcal{L}(x,u,v) \geq \min_{x} \mathcal{L}(x,u,v) =: g(u,v)$$

Lagrange dual.

dual problem: $\max_{u,v} g(u,v) = g^\star$ $\underline{\underline{u^\star, v^\star}}$

subject $u \geq 0$

$f^* \geq g^*$: weak duality

Slater's Condition

- $f, h_1 h_2 \ldots h_m$ Convex
- $l_1 \ldots l_r$ affine
- x st $h_i(x) < 0 \quad i=1, \ldots m$

  $l_j(x) = 0 \quad j=1, \ldots r$

$f^* = g^*$ strong duality

KKT Conditions

- $0 = \frac{\partial L}{\partial x}$ $\quad$ $0 = \frac{\partial L}{\partial u}$ $\quad$ $0 = \frac{\partial L}{\partial v}$ not needed.

- $u_i \cdot h_i(x) = 0 \quad i = 1, \ldots m$ [Complementary Slackness]

- $h_i(x) \leq 0 \quad l_j(x) = 0$

- $u \geq 0$

If x^*, u^*, v^* are solutions then they satisfy KKT ↑

$$\max_x -\sum_{i=1}^{k} x_i \log x_i$$

$$\text{st} \quad \sum_{i=1}^{k} x_i = 1$$

$$\min_x \quad \sum^{k} x_i \log x_i$$

$$\text{st} \quad \sum^{k} x_i = 1$$

$$\mathcal{L}(\underline{x}, \lambda) = \underline{\sum^{k} x_i \log x_i + \lambda \left(\sum x_i - 1 \right)} \qquad g(\lambda) = \underline{\min_x \mathcal{L}(\underline{x}, \lambda)}$$

$$x \log x + \lambda(x - 1) \Rightarrow \frac{x}{x} + \log x + \lambda = 0$$

$$(1 + \lambda) = -\log x$$

$$\boxed{\exp^{-(1+\lambda)} = x}$$

$$\max_{\lambda} g(\lambda)$$

$$g(\lambda) = -\sum^{k} e^{-(1+\lambda)}(1+\lambda) + \lambda\left(\sum^{k} e^{-(1+\lambda)} - 1\right)$$

$$= -k \cdot e^{-(1+\lambda)} \underline{- k\lambda e^{-(1+\lambda)}} + \underline{\lambda k e^{-(1+\lambda)}} - \lambda$$

$$= \underline{\underline{-k e^{-(1+\lambda)} - \lambda}}$$

$$\frac{d}{dx} e^{-x} = -e^{x}$$

$$+k(\cancel{1+\lambda})\, e^{-(1+\lambda)} = 1$$

$$\Rightarrow \quad k(\cancel{1+\lambda}) = e^{(1+\lambda)}$$

$$\log k \; \cancel{\log(k+1+\lambda)} = (1+\lambda) \Rightarrow \lambda = \log k - 1$$

$$x = e^{-(1+\lambda)}$$

$$\lambda = \log k - 1$$

$$= e^{-(\cancel{1} + \log k - \cancel{1})}$$

$$= e^{-\log k}$$

$$= e^{\log 1/k} = 1/k$$

$$\boxed{x_i = 1/k}$$

$$\max_{\beta, \beta_0, M} \quad 2M$$

2 classes $y_i = \{-1, 1\}$

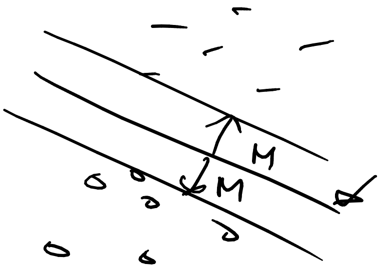
$x \in \mathbb{R}^p$

$$\text{st:} \quad y_i \left(\frac{x_i^T \beta}{\|\beta\|_2} + \frac{\beta_0}{\|\beta\|_2} \right) \geq M \quad i = 1, \dots N$$

$\{x_i, y_i\}_{i=1}^N$

Separability.

$$\boxed{M = \frac{1}{\|\beta\|_2}}$$



$$\max_{\beta, \beta_0} \frac{2}{\|\beta\|_2}$$

$$\text{st} \quad y_i(x_i^T \beta + \beta_0) \geq 1 \quad i=1, \ldots N$$

$$\min_{\beta, \beta_0} \frac{1}{2} \|\beta\|_2^2$$

$\| \ \|_2$ is Convex

$$\text{st} \quad \underline{\underline{y_i(x_i^T \beta + \beta_0) \geq 1}} \quad i=1, \ldots N$$

x^2 is Convex

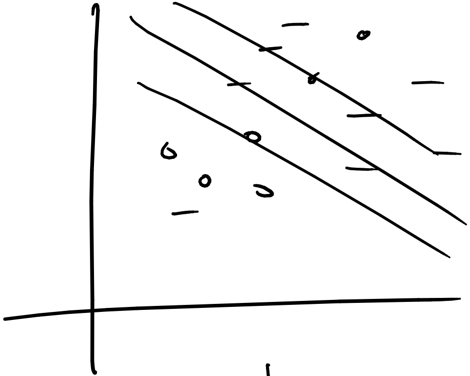
Dropping Seperability :

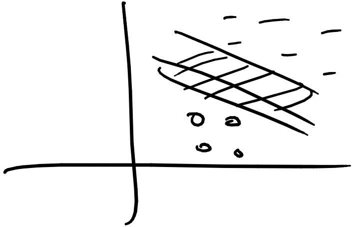
slack variables $\xi_i \quad i=1, \dots N$

$$\min_{\beta, \beta_0, \xi_i} \frac{1}{2}\|\beta\|_2^2 + C \sum_{i=1}^{N} \xi_i$$

$$\xi_i \geq 0 \quad i=1, \dots N$$

$$y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i \quad i=1, \dots N$$





$$\mathcal{L}(\beta, \beta_0, \xi_i, \underset{\geqslant 0}{\alpha_i}, \mu_i) = \frac{1}{2}\|\beta\|_2^2 + C\sum \xi_i - \sum \alpha_i\left(y_i\left(x_i^T\beta + \beta_0\right) - 1 + \xi_i\right) - \sum_{i=1}^{N} \mu_i \xi_i$$

$$\beta = \sum_{i} \alpha_i y_i x_i \qquad 0 = \sum_{i=1}^{N} \alpha_i y_i$$

$$\alpha_i = C - \mu_i \Rightarrow \mu_i = C - \alpha_i$$

$$\left[\begin{array}{l} \alpha_i \left(y_i \left(x_i^T \beta + \beta_0 \right) - 1 + \xi_i \right) = 0 \\ \mu_i \cdot \xi_i = 0 \quad \text{or} \quad (C - \alpha_i) \xi_i = 0 \end{array} \right.$$

$$\underset{\in \mathbb{R}^p}{\beta} = \sum_{i=1}^{N} y_i \alpha_i \underset{\in \mathbb{R}^p}{x_i} \longrightarrow \beta^* = \sum_{i=1}^{N} y_i \alpha_i^* x_i \quad ☆$$

$$\max_{\alpha} g(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \underbrace{x_i^T x_j}_{K(x_i, x_j)}$$

$$C \geq \alpha_i \geq 0 \quad i = 1, \dots N$$

$$0 = \sum_{i=1}^{N} \alpha_i y_i$$

☆

$$
\begin{aligned}
f(x_{test}) &= \beta^T x_{test} + \beta_0 \\
&= x_{test}^T \left(\sum_{i=1}^{N} y_i \alpha_i x_i \right) + \beta_0 \\
&= \sum_{i=1}^{N} y_i \alpha_i \underbrace{x_{test}^T x_i} + \beta_0
\end{aligned}
$$

Kernel trick to introduce nonlinearity : replace the inner products $x^T z$ with $K(x, z)$

$$K(x, z) = \underline{\underline{\phi}}(x)^T \Phi(z)$$

$$\phi(x) = \begin{bmatrix} 1 \\ x_1^2 \\ x_2^2 \\ \sqrt{2}\, x_1 \\ \sqrt{2}\, x_2 \\ \sqrt{2}\, x_1 x_2 \end{bmatrix}$$

RBF : $K(x, z) = \exp\left(-\gamma \|x - z\|_2^2\right)$

$$K(x, z) = (1 + x^T z)^2$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

$$(1 + x_1 z_1 + x_2 z_2)^2$$

$$1 + x_1^2 z_1^2 + x_2^2 z_2^2 + 2 x_1 z_1 + 2 x_2 z_2 + 2 x_1 z_1 x_2 z_2$$

$$\bar{K} = \left[\ \ \right]_{N \times N}$$

$$K(x_i, x_j)$$

$$v^T \bar{K} v \geq 0 \quad \text{for all } v \in \mathbb{R}^N$$

$N \times 1$

$$\underline{\underline{\phi(x)^T \phi(z)}}$$

$$\|Ax\|^2 \geq 0$$

① multi class

② Hinge loss

$$\min_{\beta, \beta_0} \frac{1}{2}\|\beta\|_2^2 + C \sum \xi_i$$

$$\xi_i \geq 0$$

$$y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i$$

$$\xi_i \geq 1 - y_i(x_i^T\beta + \beta_0)$$

$$\min_{\beta, \beta_0} \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{N} \max\left(0, \underbrace{1 - y_i(x_i^T\beta + \beta_0)}\right)$$

$$\left[1 - y_i(x_i^T\beta + \beta_0)\right]_+$$

Unsupervised learning

$\underline{\underline{f : X \to Y}}$ X

P_X

(a) reduce dimensionality

(b) Clustering

(c) association rules ★

AR. $\mathbb{P}_X$ $(X_1, \dots, X_p) = X$

Goal: find joint values of subsets of X that appear frequently

$X_j \in \{0,1\}$ $S_j \subseteq \{0,1\}$ $\underline{\{0\}, \{1\}, \{0,1\}}$

ϕ

find $S_1 \dots S_p$ st $\mathbb{P}\left[\bigcap_{j=1}^{p} (X_j \in S_j)\right]$ is high.

K : $\mathbb{P}\left[\prod_{k \in K} X_k = 1\right]$ is high.

$\{$ indices i where $X_i = 1 \}$

ω

itemset

frequent itemset. if $\frac{1}{N} \sum_{i=1}^{N} \prod_{i \in K} x_{ik} = \hat{\omega} > t$

support of K

Alg: $\begin{bmatrix} X_1 & \cdots & X_p \end{bmatrix}_{N \times 1} \longrightarrow \{K_j\}_{j=1}^{M}$

If $K' \subseteq K$

then $\text{support}(K') \geq \text{support}(K)$

(Apriori)

1st round: $\{1\}, \{2\}, \{4\}, \{6\}$

2nd round: $\{1,2\}, \{2,4\}$

3rd round: $\{1,2,4\}$