Tutorial 11: Given $\{(x_t, y_t)\}_{t=1}^{n}$ & $C > 0$ $x_t \in \mathbb{R}^d$

1. (P) $\min_{\theta, \theta_0, \xi} \frac{1}{2}\|\theta\|^2 + \frac{C}{2}\sum_t \xi_t^2$

(LS-SVM) $\theta \in \mathbb{R}^d$, $\theta_0 \in \mathbb{R}$: $y_t(x_t^T\theta + \theta_0) = 1 - \xi_t$ $t = 1, \ldots, n.$

$\xi \in \mathbb{R}^n$

a) Prove that the problem is equivalent to

(P') $\min_{\theta, \theta_0, \xi} \frac{1}{2}\|\theta\|^2 + \frac{C}{2}\sum_t \xi_t^2$

$x_t^T\theta + \theta_0 = y_t - \xi_t$ $t = 1, \ldots, n.$

Pf: Let $(\hat{\theta}, \hat{\theta}_0, \hat{\xi})$ be a solution to (P). It has the least cost &

$$y_t(x_t^T\hat{\theta} + \hat{\theta}_0) = 1 - \hat{\xi}_t, \quad t = 1, \ldots, n.$$

Create a new set of decision variables $(\tilde{\theta}, \tilde{\theta}_0, \tilde{\xi})$

$y_t \in \{\pm 1\}.$

$$\tilde{\theta} = \hat{\theta}, \quad \tilde{\theta}_0 = \hat{\theta}_0, \quad \tilde{\xi}_t = y_t\hat{\xi}_t, \; t = 1, \ldots, n$$

optimal for (P').

$\tilde{\xi}_t^2 = \hat{\xi}_t^2$

The $(\tilde{\theta}, \tilde{\theta}_0, \tilde{\xi})$ has the same cost as $(\hat{\theta}, \hat{\theta}_0, \hat{\xi})$.

Check constraints are satisfied $(\tilde{\theta}, \tilde{\theta}_0, \tilde{\xi})$.

$$y_t \, y_t\left(x_t^T\hat{\theta}+\hat{\theta}_0\right)=y_t(1-\hat{\xi}_t), \quad t=1,\dots,n.$$

This can be written as

$$x_t^T\tilde{\theta}+\tilde{\theta}_0 = x_t^T\hat{\theta}+\hat{\theta}_0 = y_t(1-\hat{\xi}_t) = y_t - y_t\hat{\xi}_t$$
$$= y_t - \tilde{\xi}_t.$$

The $(\tilde{\theta},\tilde{\theta}_0,\tilde{\xi})$ satisfy the constraints (P').

---

b) (P')
$$\min_{\theta,\theta_0,\xi} \frac{1}{2}\|\theta\|^2+\frac{C}{2}\sum_t \xi_t^2$$
$$x_t^T\theta+\theta_0=y_t-\xi_t \quad t=1,\dots,n.$$

Let α_t be the Lagrange mult. for the t^{th} constraint.
$\alpha=(\alpha_1,\dots,\alpha_n)$

$$L(\theta,\theta_0,\xi;\alpha)=\frac{1}{2}\|\theta\|^2+\frac{C}{2}\sum_t \xi_t^2-\sum_t \alpha_t\left(\theta^Tx_t+\theta_0-y_t+\xi_t\right)$$

$$\sum_t \alpha_t\theta_0$$

$$\nabla_\theta L(\theta,\theta_0,\xi;\alpha)= \text{①}\ \theta-\sum_t \alpha_t x_t=0$$

$$\theta=\sum_t \alpha_t x_t$$

$$\frac{\partial}{\partial\theta_0}L(\theta,\theta_0,\xi;\alpha)= \text{②}\ \sum_t \alpha_t=0 \qquad \text{(Stationarity)}$$

$$\frac{\partial}{\partial\xi_t}L(\theta,\theta_0,\xi;\alpha)= \text{③}\ C\xi_t-\alpha_t=0. \quad \Rightarrow \text{③}\ C\xi_t=\alpha_t$$

$$\text{④}\ \theta^Tx_t+\theta_0-y_t+\xi_t=0 \quad \forall t \qquad \text{(PF)}$$

$\eta=1, \quad \beta=0$

iii) $y = (y_1, \dots, y_n)^T \qquad \mathbb{1} = (1, \dots, 1)^T \in \mathbb{R}^n$

$$\begin{bmatrix} M & \mathbb{1} \\ \mathbb{1}^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \hat{\theta}_0 \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix}$$

Plugging ① & ③ into ④ yields

$$\underbrace{\left(\sum_{s=1}^{n} \alpha_s x_s \right)^T}_{\theta} x_t + \theta_0 + \frac{\alpha_t}{C} = y_t$$

By the linearity of the inner product

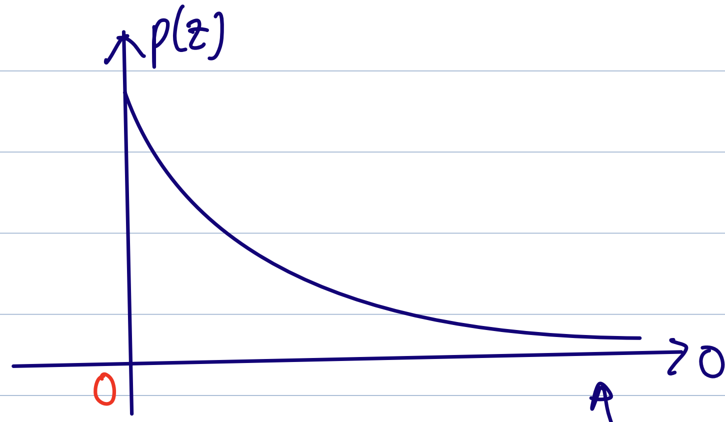
$$\sum_{s=1}^{n} \alpha_s (x_s^T x_t) + \theta_0 + \frac{\alpha_t}{C} = y_t$$

Form matrix $M \in \mathbb{R}^{n \times n}$ with (s,t) element $x_s^T x_t + \frac{\mathbb{1}\{s=t\}}{C}$

$$\begin{bmatrix} M & \mathbb{1} \\ \mathbb{1}^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \theta_0 \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix}$$

7.1 Linear measurements in exponential noise.

$$p(z) = \begin{cases} \frac{1}{a} e^{-z/a} & , \ a \geq 0 \\ 0 & , \ \text{else} \end{cases}$$



$$y_i = a_i^T x + v_i, \quad i = 1, \dots, n$$

v_i is exponentially distributed $\qquad v_i \sim p(\cdot)$

Data: $\{(y_i, a_i)\}_{i=1}^n$

Maximize the likelihood of data: $\qquad y_i - a_i^T x = v_i$

$$p_x(y) = \prod_{i=1}^n p(y_i - a_i^T x)$$

parameter

$$x = \operatorname*{argmax}_x \log p_x(y) = \operatorname*{argmax}_x \sum_{i=1}^n \log p(\underline{y_i - a_i^T x})$$

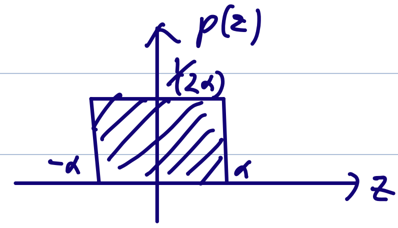
$$= \operatorname*{argmax}_x \sum_{i=1}^n \log\left(\cancel{\frac{1}{\alpha}}\, e^{-(y_i - a_i^T x)/\alpha}\, 1\{y_i \geq a_i^T x\}\right)$$

$$= \operatorname*{argmax}_x \sum_{i=1}^n -(y_i - a_i^T x)\cancel{/\alpha} \quad [\; y_i \geq a_i^T x \;]$$

MLE: $\qquad \min_x \; 1^T(y - Ax) \qquad$ s.t. $\quad y \geq Ax$

A is the matrix with a_i^T as its rows.

BV 7.2:



$$y = Ax + v$$

$$v \sim p(\cdot) \qquad p(z) = \begin{cases} \frac{1}{2\alpha} & |z| \le \alpha \\ 0 & |z| > \alpha \end{cases}$$

If α is known, MLE procedure to find x is
find x s.t. $\|Ax - y\|_\infty \le \alpha$.

Now, we don't know α & x.
MLE:

$$\underset{x,\alpha}{\operatorname{argmax}} \prod_{i=1}^{n} P_{x,\alpha}(y_i - a_i^T x).$$

$$= \underset{x,\alpha}{\operatorname{argmax}} \sum_{i=1}^{n} \log P_{x,\alpha}(y_i - a_i^T x).$$

$$= \underset{x \in \mathbb{R}^n, \alpha > 0}{\operatorname{argmax}} \sum_{i=1}^{n} \log \left[\frac{1\{|\overbrace{y_i - a_i^T x}^{=z}| \le \alpha\}}{2\alpha} \right].$$

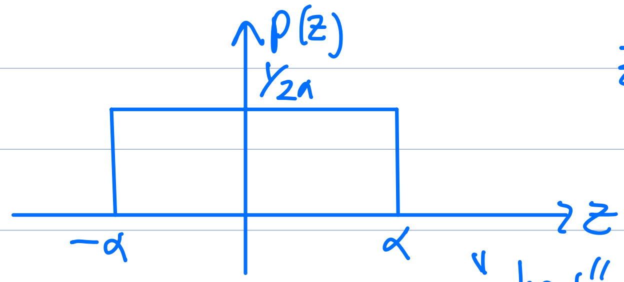
The log-likelihood function to be max over x, α

$$\ell(x,\alpha) = \begin{cases} n \log \frac{1}{2\alpha} & \|Ax - y\|_\infty \le \alpha \\ -\infty & \text{o.w.} \end{cases}$$

$\Updownarrow$

$\forall i \quad |a_i^T x - y_i| \le \alpha$

Maximize this over x & $\alpha > 0$.



$z_1, \dots, z_n$

"best" value of α

$\max_{i=1,\dots,n} |z_i|$

BV 7.3 Probit model

$y \in \{0,1\}$
$$y = \begin{cases} 1 & a^T u + b + v \le 0 \\ 0 & a^T u + b + v > 0. \end{cases}$$

$u \in \mathbb{R}^n$: explanatory variables

a, b : unknown parameters

$v \sim N(0,1)$

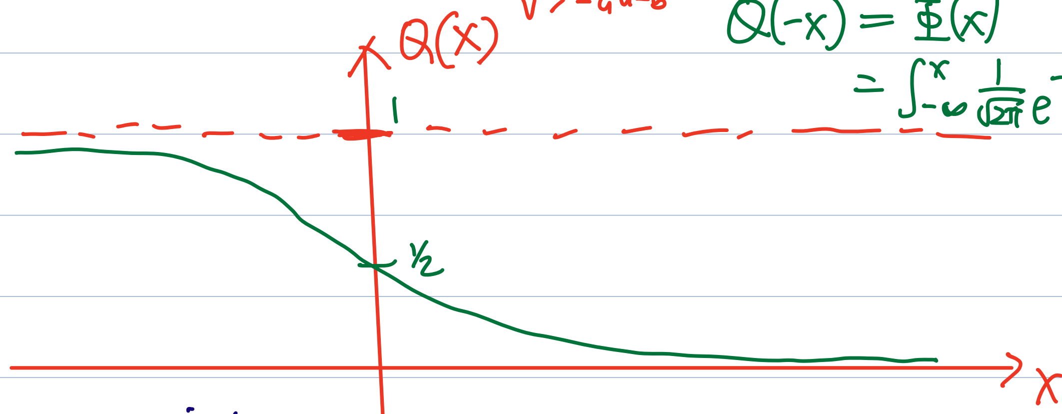
$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

$$v \le -a^T u - b$$

$$\Pr(y=1) = \Pr(a^T u + b + v \le 0) = Q(a^T u + b)$$
$$\Pr(y=0) = \Pr(a^T u + b + v > 0) = 1 - Q(a^T u + b)$$

$v > -a^T u - b$

$$Q(-x) = \Phi(x)$$
$$= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$



$i = 1, \dots, n$

(u_i, y_i) : explanatory variables & responses.

$$\operatorname*{argmax}_{a,b} \prod_{i=1}^{n} \Pr(y_i)$$

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt.$$

$$= \operatorname*{argmax}_{a,b} \sum_{i=1}^{n} \log \Pr(y_i)$$

$$= \operatorname*{argmax}_{a,b} \sum_{i:y_i=1} \log Q(\overbrace{a^T u_i + b}^{z}) + \sum_{i:y_i=0} \log(1 - Q(a^T u_i + b))$$

Is this a convex opt. problem?

Is a concave function? YES

Suffices to check that $z \mapsto \log Q(z)$ concave.