Math 156 HW 2

P1

$$P(t=C_1) = P(t=C_2) = 1/2 \qquad (\text{assume } C_1 \neq C_2)$$

$$\Rightarrow \quad [t=C_1]^c = [t=C_2] \qquad \text{(Complementary)}$$

Assumption $(*)$

$$\frac{P(x \mid t=C_1)}{P(x \mid t=C_2)} = e^{w^Tx+b} \quad \Rightarrow \quad P(x|C_1) = e^{w^Tx+b}\, P(x|C_2)$$

(1) $$P(C_2|x) = \frac{P(C_2, x)}{P(x)} = \frac{P(x|C_2)\,P(C_2)}{P(x)} = \frac{1}{2}\frac{P(x|C_2)}{P(x)}$$

$$P(x) = P(C_1 \text{ or } C_2, x) = P(C_1, x) + P(C_2, x)$$

$$= P(C_1)P(x|C_1) + P(C_2)P(x|C_2)$$

$$= \frac{1}{2} P(x|C_2)\left(1 + e^{w^Tx+b}\right)$$

$$\Rightarrow \quad P(C_2|x) = \frac{P(x|C_2)}{P(x|C_2)(1+e^{w^Tx+b})} = \frac{1}{1+e^{w^Tx+b}}$$

$$P(C_1|x) = 1 - P(C_2|x) = \frac{e^{w^Tx+b}}{1+e^{w^Tx+b}}$$

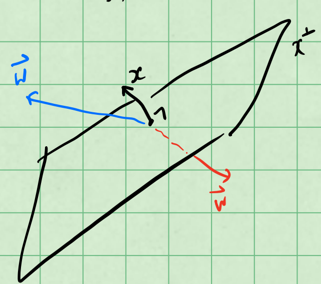
(2) $$\frac{1}{1+e^{w^Tx+b}} = \frac{e^{w^Tx+b}}{1+e^{w^Tx+b}} \iff w^Tx + b = 0$$

$$\iff x = -b + w^{\perp}$$

$\{y \in \mathbb{R}^n : w^Ty = 0\}$

$(n-1)$-dim hyperplane

(3) Note that it depends on the direction of w.



If $w^Tx > 0$, $\quad \alpha w^Tx \to \infty$ as $\alpha \to \infty$

$$\Rightarrow P(C_1|x) \to 1, \quad P(C_2|x) \to 0$$

If $w^Tx < 0$, $\quad \alpha w^Tx \to -\infty$ as $\alpha \to \infty$

$$\Rightarrow P(C_1|x) \to 0, \quad P(C_2|x) \to 1 \qquad \square$$

P2

(1) $f(x) = \log(1+e^x)$

$$f' = \frac{e^x}{1+e^x} = 1 - \frac{1}{1+e^x}$$

$$f'' = \frac{e^x}{(1+e^x)^2} > 0 \quad \forall x \in \mathbb{R}$$

(2) **Thm** if $f : \mathbb{R}^n \to \mathbb{R}$ is cvx, then

$g(x) := f(Ax+b)$ is also cvx

pf) $g(\theta x + (1-\theta)y) = f(A(\theta x + (1-\theta)y) + b) = f(\theta(Ax+b) + (1-\theta)(Ay+b))$

$\leq \theta f(Ax+b) + (1-\theta) f(Ay+b) = \theta g(x) + (1-\theta) g(y)$ □

$L(w) = f \circ \ell(w)$ where $\ell(w) = -t \tilde{x} w$, a linear map.

(3) $$\int_u^v \frac{e^x}{(1+e^x)^2}\,dx \leq \max_{x\in\mathbb{R}} \frac{e^x}{(1+e^x)^2} |v-u|$$

Set $t = 1+e^x \geq 1$. Want to Compute $\max_{t\geq 1} \frac{t-1}{t^2}$

i.e., find smallest α that makes (*) holds $\forall t \geq 1$:

(*) $t-1 \leq \alpha t^2 \iff \alpha(t - \frac{1}{2\alpha})^2 + (1 - \frac{1}{4\alpha}) \geq 0$

$\Rightarrow$ when $1 - \frac{1}{4\alpha} \geq 0$, (*) holds for all $t \in \mathbb{R}$

$\Updownarrow$

$\alpha \geq 1/4$.

Q: Is $\alpha = \frac{1}{4}$ optimal? (i.e., is smaller α possible?)

A: Yes Take $t = 2$, or, $x = 0$

then $\frac{e^x}{(1+e^x)^2} = \frac{1}{4}$.

Thus, f is $1/4$ - smooth. □

P3

(1) $A = a, \ B = b, \ C = c, \ D = d$ then

$$M = (A - BD^{-1}C)^{-1} = \left(a - \frac{bc}{d}\right)^{-1} = \frac{d}{ad-bc}$$

$$(\text{formula}) = \begin{bmatrix} \frac{d}{ad-bc} & -\frac{db}{(ad-bc)d} \\ -\frac{cd}{d(ad-bc)} & \frac{1}{d} + \frac{cdb}{d^2(ad-bc)} \end{bmatrix}$$

$$= \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & (**) \end{bmatrix}$$

$$(**) = \frac{ad-bc}{d} + \frac{bc}{d} = a$$

(2)

$$X = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

blocks ↓

$$X_{11} = MA - MBD^{-1}C = M(A - BD^{-1}C) = I$$

$$X_{12} = MB - MBD^{-1}D = MB - MB = O$$

$$X_{21} = -D^{-1}CMA + D^{-1}C + D^{-1}CMBD^{-1}C$$

$$= -D^{-1}C(\underline{MA} - I - \underline{MBD^{-1}C}) = O$$

$$X_{22} = \underline{-D^{-1}CMB} + D^{-1}D + \underline{D^{-1}CMBD^{-1}D} = I$$

P4 Set $X = X_i$ with prob. λ_i

then $$P(X \le c) = \sum_{i=1}^{k} P(X \le c \mid X = X_i)\, P(X = X_i)$$

$$= \sum_{i=1}^{k} \lambda_i\, P(X_i \le c) = \sum_{i=1}^{k} \lambda_i \int^{c} p_i(x)\,dx$$

Thus this X has a pdf $\Sigma \lambda_i p_i$

$$Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

$$= \sum_{i=1}^{k} \mathbb{E}((X - \bar{\mu})^2 \mid X = X_i)\, P(X = X_i)$$

$$= \sum_{i=1}^{k} \lambda_i\, \mathbb{E}((X_i - \bar{\mu})^2) \quad \text{——} \quad (*)$$

$$\mathbb{E}((X_i - \bar{\mu})^2) = \mathbb{E}((X_i - \mu_i + \mu_i - \bar{\mu})^2)$$

$$= \mathbb{E}((X_i - \mu_i)^2) + 2\,\mathbb{E}((X_i - \mu_i))(\mu_i - \bar{\mu}) + (\mu_i - \bar{\mu})^2$$

$$= \sigma_i^2 + (\mu_i - \bar{\mu})^2.$$

Plugging this into $(*)$, we get the desired result $\square$