ML HW-2

Q1) Distance Between 2 hyperplanes (10 points)

A1) Given that $a^T x = b$.

Let x_1 & x_2 be 2 points on planes P_1 & P_2.

$$a^T x_1 = b_1$$
$$a^T x_2 = b_2$$

$$\min_{x_1, x_2} \text{distance} \quad \|x_1 - x_2\|^2$$
subject to
$$a^T x_1 = b_1, a^T x_2 = b_2 \quad \hookrightarrow (1)$$

$$\Rightarrow a^T(x_1 - x_2) = b_1 - b_2$$

By using Cauchy-Schwarz Inequality

$$|a^T(x_1 - x_2)| \leq \|a\| \, \|x_1 - x_2\|$$

$$\Rightarrow \|x_1 - x_2\| \leq \frac{|b_1 - b_2|}{\|a\|}$$

$\rightarrow$ This is satisfied for any x_1 & x_2 for $a^T x_1 = b_1$ & $a^T x_2 = b_2$

$\Rightarrow$ Also, x_1 & x_2 can be represented as:

$$x_1 = \left(\frac{b_1}{\|a\|}\right) a \quad \& \quad x_2 = \left(\frac{b_2}{\|a\|}\right) a$$

$$\Rightarrow \|x_1 - x_2\| = \frac{|b_1 - b_2|}{\|a\|} \rightarrow (2)$$

$\Rightarrow$ Here (2) is the sol^{n} to (1) & the distance is given by:

$$\text{Dist}(x, y) = \frac{|b_1 - b_2|}{\|a\|}$$

(Q2) Determine if given functions are convex, concave, or neither. (10 points)

A2 a $E[x] = \sum_{i=1}^{K} x_i \, P[x = a_i]$

$$E[x] = \sum_{i=1}^{K} a_i * p_i$$

$$E[x] = a^T p$$

where $a = (a_1, a_2, \dots, a_K)$

$\Rightarrow$ This is a linear function of "p" and can be both concave & convex.

b $P[x > \alpha] = \sum_{i, a_i > \alpha} p_i x_i$

$$P[x > \alpha] = z^T p$$

where $z = \begin{cases} 0 & a_i \leq \alpha \\ 1 & a_i > \alpha \end{cases}$

$\Rightarrow$ This is a linear function of "p" and can be both concave & convex

c $P[\alpha < x < \beta] = \sum_{i, \alpha < a_i < \beta} p_i x_i$

$$P[\alpha < x < \beta] = C^T p$$

wher $C = \begin{cases} 1 & \alpha < a_i < \beta \\ 0 & \text{otherwise} \end{cases}$

$\Rightarrow$ This is a linear function of "p" and canbe both concave & convex

(d) $-\sum_{i=1}^{K} p_i \log(p_i)$

$\Rightarrow$ Since $a\log(a)$ is a convex function of "a", $-\sum_{i=1}^{K} p_i \log(p_i)$ is a concave function of "p".

(e) $\operatorname{Var}(x) = E[x^2] - (E[x])^2$

$$\operatorname{Var}(x) = \sum_{i=1}^{K} a_i^2 p_i - (a^T p)^2$$

$\Rightarrow$ This is a quadratic function with the Hessian Matrix $-aa^T$, which is negative semidefinite. $\therefore$ It is a concave function of "p".

Q3 Log-Concavity of Gaussian cummulative distribution function.

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$$

(a) Prove $\phi''(t)\ \phi(t) \leq (\phi'(t))^2$ for $t \geq 0$

$$\Rightarrow \phi'(t) = e^{-t^2/2} * \frac{1}{\sqrt{2\pi}}$$

$$\Rightarrow \phi''(t) = \frac{-t\, e^{-t^2/2}}{\sqrt{2\pi}}$$

→ If $\phi''(t) \leq 0$ for $t \geq 0$ &

RHS is always +ve for $t \geq 0$

⇒ Hence,

$$\phi''(t)\,\phi(t) \leq (\phi'(t))^2$$

(6) Prove $\frac{x^2}{2} \geq \frac{-t^2}{2} + tx$

Since $\frac{x^2}{2}$ is convex function, then we use first order taylor approximation which holds for differentiable convex function.

$$g(x) \geq g(t) + g'(t)(x-t)$$

$$[\text{Here } g(x) = x^2/2]$$

$$\Rightarrow x^2/2 \geq t^2/2 + t(x-t)$$

$$\Rightarrow x^2/2 \geq t^2/2 + tx - t^2$$

$$\Rightarrow x^2/2 \geq \frac{-t^2}{2} + tx$$

Hence proved

ⓒ Since we have $\frac{x^2}{2} \geq \frac{-t^2}{2} + tx$

$$\Rightarrow -\frac{x^2}{2} \leq \frac{t^2}{2} - tx$$

→ Taking exponential on both sides of inequality,

$$e^{-x^2/2} \leq e^{(t^2/2 - tx)}$$

→ Integrate it over "t",

$$\int_{-\infty}^{t} e^{-x^2/2} dx \leq \int_{-\infty}^{t} e^{(t^2/2 - tx)} dx$$

$$\Rightarrow \int_{-\infty}^{t} e^{-x^2/2} dx \leq e^{t^2/2} \int_{-\infty}^{t} e^{-tx} dx$$

(d) $\phi''(t)\,\phi(t) \le (\phi'(t))^2$ for $t<0$

Putting value of $\phi''(t)$ & $\phi'(t)$ from (a),

$$\frac{-t\,e^{-t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{t} \frac{e^{-x^2/2}\,dt}{\sqrt{2\pi}} \le \left(\frac{e^{-t^2/2}}{\sqrt{2\pi}}\right)^2$$

$$\Rightarrow -t\,e^{-t^2/2} \int_{-\infty}^{t} e^{-x^2/2}\,dt \le e^{-t^2}$$

→ Since $t<0$, $-t$ is +ve

$$\int_{-\infty}^{t} e^{-x^2/2}\,dt \le \frac{e^{-t^2/2}}{-t}$$

$$\left[\because \int_{-\infty}^{t} e^{-xt}\,dt = \frac{e^{-t^2}}{-t}\right]$$

(Q4) Equivalence of 2 convex problems.

(A4) $\Rightarrow$ Let θ be fixed for problem (b). We see at optimum of (b) we must have $u_i + v_i = |\phi_i^T\theta - \psi_i|$, because otherwise we can further decrease the objective function without violating the constraints.

$$\Rightarrow v_i = |\phi_i^T\theta - \psi_i| - u_i$$

at optimum

$\rightarrow$ Eliminating v yields the following problem.

$$\underset{u}{\text{minimize}} \sum_{i=1}^{n} \left(u_i^2 - 2Mu_i + 2M|\phi_i^T\theta - \psi_i|\right)$$

$$\text{subject to } 0 \le u_i \le \min(M, |\phi_i^T\theta - \psi_i|)$$
$$i = 1, \dots, n$$

→ The problem is separable over each u_i & we rewrite it as

$$\underset{u_i}{\text{minimize}} (u_i - M)^2 - M^2 + 2M|\phi_i^T\theta - \psi_i|$$

$$\text{subject to } 0 \le u_i \le \min(M, |\phi_i^T\theta - \psi_i|)$$

→ Here if $M < |\phi_i^T\theta - \psi_i|$, then we should choose $u_i = M$ to minimize the objective, else choose u_i to be as close to M as possible, which is $|\phi_i^T\theta - \psi_i|$. Thus, for fixed θ in problem (b), the optimal value is given by the Huber function

$$\sum_{i=1}^{n} h(\phi_i^T \theta - \psi_i)$$

Q5) Wine Quality Regression Models. [30 points]

A5)
⇒ The code is attached to the "ML_HW2.ipynb" file.

⇒ MAE for each types of loss function:

a) Least Square Loss - 53.296%

b) Huber Loss - 53.112%

c) Hinge Loss - 54.729%

Q6 Ionosphere Classification Models. [30 points]

A6 ⇒ The code is attached to the "ML_HW2.ipynb" file.

⇒ Accuracy for each types of loss function:

a) Least Square Loss - 54.901%
b) Logistic Loss - 96.078%
c) Hinge Loss - 100.000%

⇒ This value might be because the test dataset only consists of the "good" values.

X — X — X — X