# 人工智慧導論 HW2

許博翔

May 22, 2024

**Problem 1.** $\text{err}(\mathbf{w}^T\mathbf{x}, \mathbf{y}) = \max(1 - y\mathbf{w}^T\mathbf{x})^2$

$= \begin{cases} (1-\mathbf{y}\mathbf{w}^T\mathbf{x})^2, \text{ if } y\mathbf{w}^T\mathbf{x} \le 1 \\ 0, \text{ otherwise} \end{cases}$.

$\Rightarrow \nabla(\text{err}(\mathbf{w}^T\mathbf{x}, \mathbf{y})) = \begin{cases} 2(1-\mathbf{y}\mathbf{w}^T\mathbf{x}) \times \nabla(1-\mathbf{y}\mathbf{w}^T\mathbf{x}) = 2(1-\mathbf{y}\mathbf{w}^T\mathbf{x})(-\mathbf{y}\mathbf{x}) = 2(\mathbf{w}^T\mathbf{x}-\mathbf{y})\mathbf{x}, \text{ if } y\mathbf{w}^T\mathbf{x} \le \\ 0, \text{ otherwise} \end{cases}$

$\therefore \nabla \text{E}_{\text{in}}(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N} \nabla \text{err}(\mathbf{w}^T\mathbf{x}, \mathbf{y}) = \frac{1}{N}\sum_{1\le n\le N, y_n\mathbf{w}^T\mathbf{x}_n \le 1} 2(\mathbf{w}^T\mathbf{x}_n - \mathbf{y}_n)\mathbf{x}_n + \frac{1}{N}\sum_{1\le n\le N, y_n\mathbf{w}^T\mathbf{x}_n > 1} 0$

$= \frac{2}{N}\sum_{1\le n\le N, y_n\mathbf{w}^T\mathbf{x}_n \le 1} (\mathbf{w}^T\mathbf{x}_n - \mathbf{y}_n)\mathbf{x}_n$.

**Problem 2.** Let $f_u$ be the pdf of $\mathcal{N}(u, 1)$.

**Lemma 2.1.** The maximizer of $g(u) := \prod_{n=1}^{N} f_u(x_n)$ is $u = \sum_{n=1}^{N} x_n$.

*Proof.* $f_u(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-u)^2}$.

$\Rightarrow g(u) = \prod_{n=1}^{N} f_u(x_n) = (\frac{1}{\sqrt{2\pi}})^N e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-u)^2}$.

$g'(u) = (\frac{1}{\sqrt{2\pi}})^N e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-u)^2} \times \frac{d(-\frac{1}{2}\sum_{n=1}^{N}(x_n-u)^2)}{du} = (\frac{1}{\sqrt{2\pi}})^N e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-u)^2} \times (-\frac{1}{2}(\sum_{n=1}^{N} 2(u-$
$x_n))) = (\frac{1}{\sqrt{2\pi}})^N e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-u)^2} \times \sum_{n=1}^{N}(x_n - u)$.

$\because (\frac{1}{\sqrt{2\pi}})^N e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-u)^2} > 0$ for all $u$.

$\therefore g'(u) = 0 \iff \sum_{n=1}^{N}(x_n - u) = 0 \iff u = \frac{1}{N}\sum_{n=1}^{N} x_n.$

$\therefore u = \frac{1}{N}\sum_{n=1}^{N} x_n$ is the minimizer of $g(u)$. ■

$\because \mathcal{N}(\mathbf{u}, I)$ is the joint distribution of $d$ independent random variables $X_1, X_2, \ldots, X_d$, where $X_i \sim \mathcal{N}(\mathbf{u}_i, 1)$.

$\therefore p_{\mathbf{u}}(\mathbf{x}_n) = \prod_{i=1}^{d} f_{\mathbf{u}_i}(\mathbf{x}_{n,i})$.

$\Rightarrow \prod_{n=1}^{N} p_{\mathbf{u}}(\mathbf{x}_n) = \prod_{n=1}^{N}\prod_{i=1}^{d} f_{\mathbf{u}_i}(\mathbf{x}_{n,i}) = \prod_{i=1}^{d}\prod_{n=1}^{N} f_{\mathbf{u}_i}(\mathbf{x}_{n,i}) \overset{\text{By } \textbf{Lemma (2.1)}}{\leq} \prod_{i=1}^{d}\prod_{n=1}^{N} f_{\frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_{j,i}}(\mathbf{x}_{n,i}) = \prod_{i=1}^{d}\prod_{n=1}^{N} f_{\mathbf{u}_i^*}(\mathbf{x}_{n,i}) = \prod_{n=1}^{N}\prod_{i=1}^{d} f_{\mathbf{u}_i^*}(\mathbf{x}_{n,i}) = \prod_{n=1}^{N} p_{\mathbf{u}^*}(\mathbf{x}_n).$

$\therefore \mathbf{u}^*$ is the maximizer of $\prod_{n=1}^{N} p_{\mathbf{u}}(\mathbf{x}_n)$.

**Problem 3.**

$\mathbf{z}_1 = [1, 1, 1, 1, 1, 1]$.

$\mathbf{z}_2 = [1, -1, 1, 1, -1, 1]$.

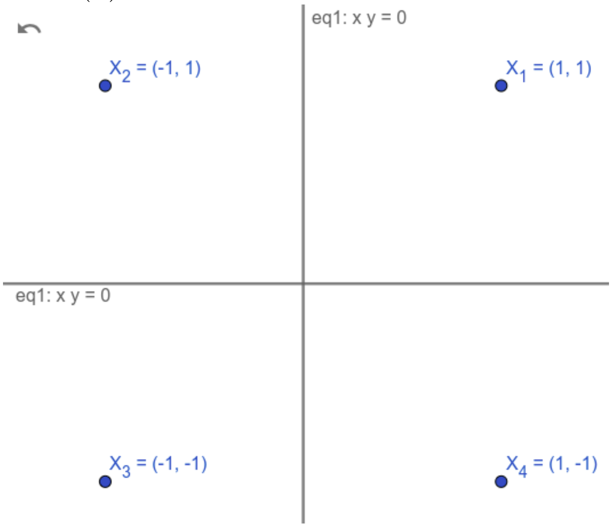
$\mathbf{z}_3 = [1, -1, -1, 1, 1, 1]$.

$\mathbf{z}_4 = [1, 1, -1, 1, -1, 1]$.

One can see that $\mathbf{z}_{n,5} = -y_n$ for all $n = 1, 2, 3, 4$.

$\therefore \tilde{\mathbf{w}} = [0, 0, 0, 0, -1, 0]$ satisfies the condition.

$\tilde{\mathbf{w}}^T \Phi_2(\mathbf{x}) = 0$ is the curve $-x_1x_2 = 0$.



**Problem 4.** Let $A = \sum_{1\le n\le N, g_t(\mathbf{x}_n)\ne y_n} w_n^t, B = \sum_{1\le n\le N, g_t(\mathbf{x}_n)= y_n} w_n^t$.

$\Rightarrow \epsilon_t = \dfrac{A}{A+B}$.

$d_t = \sqrt{\dfrac{1-\epsilon_t}{\epsilon_t}} = \sqrt{\dfrac{B}{A}}$.

$\sum_{n=1}^{N} w_n^{t+1}\delta(g_t(\mathbf{x}_n), y_n) = \sum_{1\le n\le N, g_t(\mathbf{x}_n)\ne y_n} w_n^{t+1} = d_t \sum_{1\le n\le N, g_t(\mathbf{x}_n)\ne y_n} w_n^t = d_t A = \sqrt{\dfrac{B}{A}}A = \sqrt{AB}$.

$\sum_{n=1}^{N} w_n^{t+1} = \sum_{1\le n\le N, g_t(\mathbf{x}_n)\ne y_n} w_n^{t+1} + \sum_{1\le n\le N, g_t(\mathbf{x}_n)= y_n} w_n^{t+1} = \sum_{1\le n\le N, g_t(\mathbf{x}_n)\ne y_n} d_t w_n^t + \sum_{1\le n\le N, g_t(\mathbf{x}_n)= y_n} w_n^t \dfrac{1}{d_t} = Ad_t + \dfrac{B}{d_t} = A\sqrt{\dfrac{B}{A}} + B\sqrt{\dfrac{A}{B}} = 2\sqrt{AB}$.

$\therefore \dfrac{\sum_{n=1}^{N} w_n^{t+1}\delta(g_t(\mathbf{x}_n), y_n)}{\sum_{n=1}^{N} w_n^{t+1}} = \dfrac{\sqrt{AB}}{2\sqrt{AB}} = 0.5$.