Homework 4 Solutions

Problem 1: We did this in the lecture, see 4p1.

Problem 2:

$$\mathcal{L}(\theta) = \prod_{n=1}^{N} \prod_{c=1}^{6} \theta_c^{\mathbb{I}(y=c)}$$

note $\theta_6 = 1 - \theta_1 - \ldots - \theta_5$
so only five unknowns

$$\begin{aligned}
NLL(\theta) &= -\ln \mathcal{L}(\theta) \\
&= -\sum_{n=1}^{N} \sum_{c=1}^{6} \mathbb{I}(y=c) \cdot \ln(\theta_c) \\
&= -\sum_{n: y_n=1} \ln(\theta_1) - \ldots - \sum_{n: y_n=6} \ln(\theta_6) \\
&= -N_1 \ln(\theta_1) - \ldots - N_5 \cdot \ln(\theta_5) - N_6 \cdot \ln(1 - \theta_1 - \ldots - \theta_5)
\end{aligned}$$

$$\frac{\partial}{\partial \theta_k} NLL(\theta) = -\frac{N_k}{\theta_k} + \frac{N_6}{1 - \sum_{\ell=1}^{5} \theta_\ell} = 0 \qquad k = 1..5$$

$$\Rightarrow \frac{\theta_k}{N_k} = \frac{1 - \sum_{\ell=1}^{5} \theta_\ell}{N_6} \quad \Rightarrow \quad \theta_k + N_k \sum_{\ell=1}^{5} \theta_\ell = \frac{N_k}{N_6}$$

this is a linear system. The matrix form is

$$\left(I + \frac{1}{N_6} \underbrace{\begin{bmatrix} N_1 \\ \vdots \\ N_5 \end{bmatrix}}_{u} \cdot \underbrace{[1,1,1,1,1]}_{v^T} \right) \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_5 \end{bmatrix} = \frac{1}{N_6} \underbrace{\begin{bmatrix} N_1 \\ \vdots \\ N_5 \end{bmatrix}}_{u}$$

By the Sherman-Morrison formula $(I + \underline{u}\underline{v}^T)^{-1} = I - \gamma \underline{u}\underline{v}^T$ where $\gamma = \frac{1}{1 + v^T u}$

It follows that

$$\gamma = \frac{1}{1 + \frac{N_1 + \ldots + N_5}{N_6}} = \frac{N_6}{N}$$

$$\begin{bmatrix} \theta_1 \\ \vdots \\ \theta_5 \end{bmatrix} = \left(I - \underbrace{\frac{N_6}{N}}_{\gamma} \frac{1}{N_6} \underbrace{\begin{bmatrix} N_1 \\ \vdots \\ N_5 \end{bmatrix}}_{u} \cdot [1, \ldots, 1] \right) \cdot \frac{1}{N_6} \begin{bmatrix} N_1 \\ \vdots \\ N_5 \end{bmatrix}$$

$$\Rightarrow \theta_k = \frac{N_k}{N_6} - \frac{1}{N} N_k \cdot \sum_{\ell=1}^{5} N_\ell \cdot \frac{1}{N_6}$$

$$= \frac{N_k}{N_6} \cdot \left(1 - \frac{1}{N} \sum_{\ell=1}^{5}\right) = \frac{N_k}{N_6} \cdot \frac{1}{N} \underbrace{\left(N - \sum_{\ell=1}^{5} N_\ell\right)}_{= N_6} = \frac{N_k}{N}$$

Note: In sec. 4.2.4 the method of Lagrange multipliers is used to find the optimal $\theta_1 \ldots \theta_6$ under the constraint $\sum \theta_k = 1$.
Both methods are good ways to solve this problem.

HW4 p2

Problem 3 $p(X=x \mid Y=1)\,\pi_1 = p(X=x \mid Y=2)\,\pi_2$

$$\Leftrightarrow \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)\pi_1 = \frac{1}{\sqrt{2\pi}\,\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)\pi_2$$

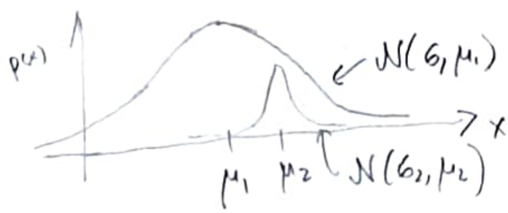
$$\Leftrightarrow \ln\frac{\pi_1}{\sigma_1} - \left(\frac{x-\mu_1}{\sigma_1}\right)^2 = \ln\frac{\pi_2}{\sigma_2} - \left(\frac{x-\mu_2}{\sigma_2}\right)^2$$

$$\Leftrightarrow \frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 - \frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \ln\frac{\pi_2}{\sigma_2} + \ln\frac{\pi_1}{\sigma_1} = 0$$

$$\Leftrightarrow \frac{1}{2}\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\right)^2 x^2 + bx + c = 0$$

can have zero one or two solutions

zero solutions:



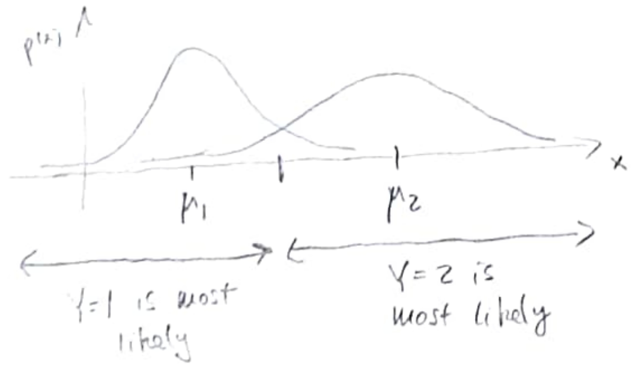
μ_1 & μ_2 are close

$\pi_2 \mathcal{N}(\sigma_2, \mu_2)$ has a small amplitude

$\pi_1 \mathcal{N}(\sigma_1, \mu_1)$ " " larger amplitude

In this case the most likely class is always $Y=1$.

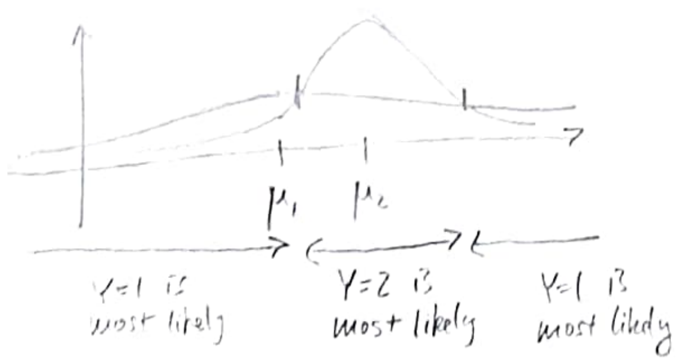
one solution:



μ_1 and μ_2 are separated

or: $\sigma_1 = \sigma_2$

two solutions



$\sigma_1 \ll \sigma_2$

μ_1 & μ_2 are close

Problem 4

$$p(Y=1) = \sigma(b + w_1 x_1 + w_2 x_2) \qquad x_1 = \text{hrs} \quad w_1 = 0.05 \qquad b = -6$$
$$x_2 = \text{gpa} \quad w_2 = 1$$

a.) $\sigma(-6 + \underbrace{\overset{\frac{1}{20}}{0.05}(40)}_{2} + 1 \cdot 3.5) = \sigma(-0.5) = \frac{1}{1+e^{0.5}} \approx 37.6\%$

b.) Note: $\sigma(0) = \frac{1}{2} \Rightarrow$ increase x_1 such that

$$-6 + \frac{1}{20} \cdot x_1 + 3.5 \geq 0$$
$$\Rightarrow x_1 \geq 20 \cdot 2.5 = 50.$$

Problem 5 $\quad Y \in \{0, 1\}$ No / Yes $\qquad X =$ last year's profits

$$p(Y=0 \mid X=x) = \frac{1}{\sqrt{2\pi} \cdot 6} \cdot \exp\left(-\frac{x^2}{2 \cdot 36}\right) \qquad \pi_0 = 0.2$$

$$p(Y=1 \mid X=x) = \frac{1}{\sqrt{2\pi} \cdot 6} \cdot \exp\left(-\frac{(x-10)^2}{2 \cdot 36}\right) \qquad \pi_1 = 0.8$$

cancels

Baye's Law

$$p(X=x \mid Y=1) = \frac{\exp\left(-\frac{(x-10)^2}{72}\right) \cdot 0.8}{\exp\left(-\frac{x^2}{72}\right) 0.2 + \exp\left(-\frac{(x-10)^2}{72}\right) \cdot 0.8}$$

$$p(X=4 \mid Y=1) = \frac{\exp\left(-\frac{36}{72}\right)}{\exp\left(-\frac{16}{72}\right) 0.2 + \exp\left(-\frac{36}{72}\right) \cdot 0.8} = \frac{\exp\left(-\frac{1}{2}\right)}{\exp\left(-\frac{2}{9}\right) 0.2 + \exp\left(-\frac{1}{2}\right) 0.8}$$

$$\approx 94.0\%$$