(1

$\hat{\mu} = \frac{1}{n} \Sigma x_i \qquad \hat{\Sigma} = \frac{1}{n} \Sigma (x_i - \hat{\mu})(x_i - \hat{\mu})^T$

$A: \hat{\mu}_A = \frac{1}{7}\left(\begin{bmatrix} -2 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 5 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 1.85 \\ 2 \end{bmatrix}$

$\Rightarrow \hat{\Sigma}_A = \begin{bmatrix} 5.83 & 2.28 \\ 2.28 & 2.28 \end{bmatrix}$

$\hat{\mu}_B = \frac{1}{5}\left(\begin{bmatrix} 0 \\ -4 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \end{bmatrix} + \begin{bmatrix} -2 \\ -1 \end{bmatrix} + \begin{bmatrix} -3 \\ 1 \end{bmatrix} + \begin{bmatrix} -4 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -2.2 \\ -1.2 \end{bmatrix} \qquad \Rightarrow \hat{\Sigma}_B = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 1.66 \end{bmatrix}$

با توجه به اینکه اطلاعات خاصی از احتمال پیشین داده نشده است، ملاک آن را تعداد نقاط قرار می دهیم:

$P(A) = \frac{7}{12} \qquad P(B) = \frac{5}{12}$

$g_i(x) = x^T W_i x + w_i^T x + w_{i0} \qquad W_i = -\frac{1}{2} \Sigma_i^{-1} \qquad w_i = \Sigma_i^{-1} \mu_i$

$w_{i0} = -\frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln|\Sigma_i| + \ln[P(w_i)]$

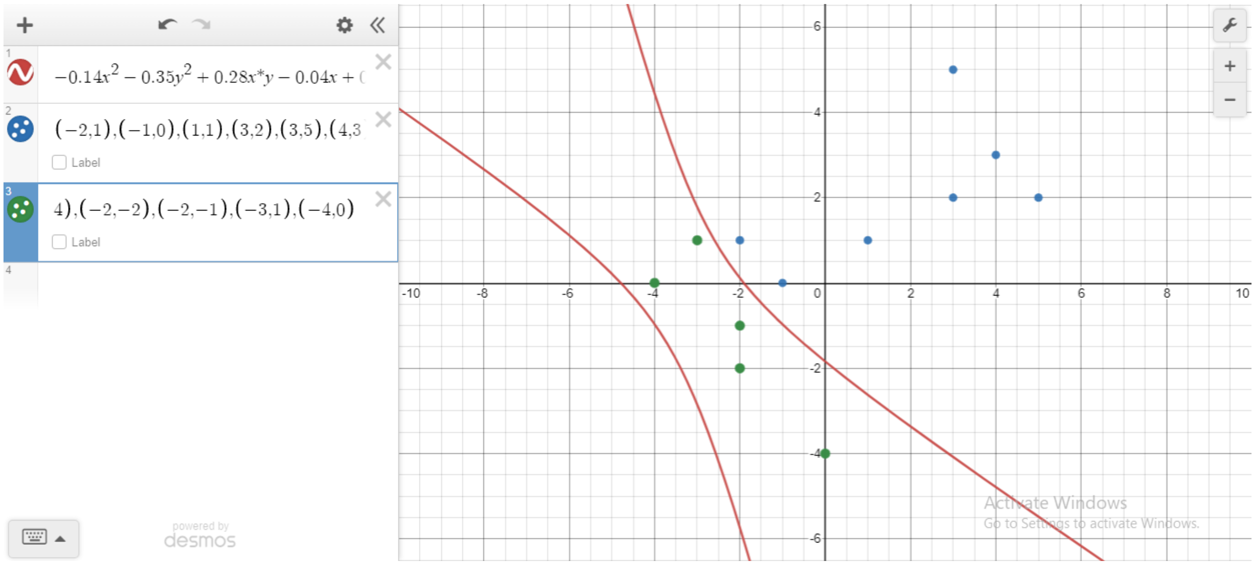
$\Rightarrow g_A: W_A = \begin{bmatrix} -0.14 & 0.14 \\ 0.14 & -0.36 \end{bmatrix} \qquad w_A = \begin{bmatrix} -0.04 \\ 0.91 \end{bmatrix} \qquad w_{A0} = -2.46$

$g_B: W_B = \begin{bmatrix} -1.41 & -0.97 \\ -0.97 & -0.84 \end{bmatrix} \qquad w_B = \begin{bmatrix} -8.55 \\ -6.29 \end{bmatrix} \qquad w_{B0} = -14.08$

با برابری مقادیر $g_A = g_B$: $\Rightarrow -0.14 x_1^2 - 0.35 x_2^2 + 0.28 x_1 x_2 - 0.04 x_1 + 0.915 x_2 - 2.46 =$

$-1.41 x_1^2 - 0.84 x_2^2 - 1.95 x_1 x_2 - 8.55 x_1 - 6.3 x_2 - 14.08$

$\Rightarrow \boxed{1.27 x_1^2 + 2.23 x_1 x_2 + 8.51 x_1 + 0.49 x_2^2 + 7.215 x_2 + 11.62 = 0}$


−0.14x^2 − 0.35y^2 + 0.28x*y − 0.04x + (
(−2,1),(−1,0),(1,1),(3,2),(3,5),(4,3
Label
4),(−2,−2),(−2,−1),(−3,1),(−4,0)
Label
powered by
desmos
-10 -8 -6 -4 -2 0 2 4 6 8 10
6 4 2 -2 -4 -6
Activate Windows
Go to Settings to activate Windows.

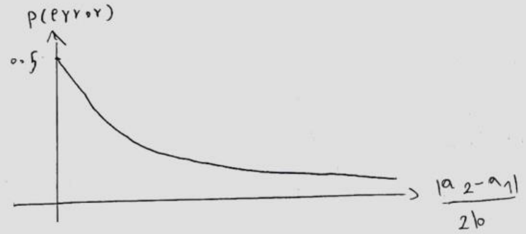
2) الف) فرض می‌کنیم $a_2 > a_1$ ، تفاوتی در حل مساله و کلیت ایجاد نمی‌کند. مرز تصمیم در $\frac{a_1+a_2}{2}$ قرار دارد ؛

$$p(error) = \int_{-\infty}^{\frac{a_1+a_2}{2}} p(w_2|x)\,dx + \int_{\frac{a_1+a_2}{2}}^{\infty} p(w_1|x)\,dx = \frac{1}{\pi b}\int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{0.5}{1+\left(\frac{x-a_2}{b}\right)^2}\,dx + \frac{1}{\pi b}\int_{\frac{a_1+a_2}{2}}^{\infty} \frac{0.5}{1+\left(\frac{x-a_1}{b}\right)^2}\,dx$$

$$= \frac{1}{\pi b}\int_{-\infty}^{\frac{a_1-a_2}{2}} \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}\,dx = \frac{1}{\pi}\int_{-\infty}^{\frac{a_1-a_2}{2}} \frac{1}{1+y^2}\,dy \rightarrow \frac{x-a_2}{b} = y$$ تغییر متغیر

$$\Rightarrow p(error) = \frac{1}{\pi}\left[\tan^{-1}\left|\frac{a_1-a_2}{b}\right| - \tan^{-1}[-\infty]\right] = \boxed{\frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2-a_1}{2b}\right|}$$

ب)



ج)

$$p(error) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2-a_1}{2b}\right| \qquad -\frac{\pi}{2} \leq \tan^{-1} \leq \frac{\pi}{2} \Rightarrow \qquad 0 \leq p(error) \leq \frac{1}{2}$$

$$\boxed{\begin{matrix} a_2 = a_1 \\ \text{یا} \\ b \to \infty \end{matrix}} \Leftarrow \tan^{-1}\left|\frac{a_2-a_1}{2b}\right| = \frac{\pi}{2} \Leftarrow \max(p(error)) = \frac{1}{2} \Leftarrow$$

---

3) الف)

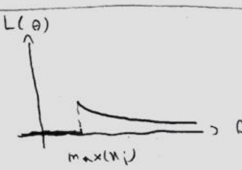
$$\text{log-liklihood} \propto \sum x_i \log p_i \Rightarrow \max \sum x_i \log p_i \quad s.t.\ \sum p_i = 1$$

$$\mathcal{L} = \sum x_i \log p_i + \lambda\left(\sum p_i - 1\right) \Rightarrow \frac{d\mathcal{L}}{dp_i} = 0 \Rightarrow \frac{x_i}{p_i} + \lambda = 0 \Rightarrow x_i = -\lambda p_i$$

$$\Rightarrow \underbrace{\sum x_i}_{n} = -\lambda \underbrace{\sum p_i}_{1} = -\lambda \Rightarrow \lambda = -n \Rightarrow \boxed{p_i^* = \frac{x_i}{n}}$$

ب)

$$\text{liklihood} = \begin{cases} 0 & \text{اگر یکی از } x_i \text{ ها بزرگتر از } \theta \text{ باشد} \\ \left(\frac{1}{\theta}\right)^n & o.w. \end{cases} \equiv \begin{cases} 0 & o.w. \\ \left(\frac{1}{\theta}\right)^n & \max(x_i) \leq \theta \end{cases}$$



$\left(\frac{1}{\theta}\right)^n$ تابع نزولی از θ می‌باشد $\Rightarrow \boxed{\theta^* = \max(x_i)}$

$\bar{V} \overset{\Delta}{=} \max x_i \qquad F_{\bar{V}}(v) = P(\bar{V} \leqslant v) = P(\max x_i \leqslant x) = P(x_1 \leqslant v, x_2 \leqslant v, \dots, x_n \leqslant v)$ (۲

$\overset{iid}{=} P(x_1 \leqslant v)\, P(x_2 \leqslant v) \cdots P(x_n \leqslant v) = \left(\frac{v}{\theta}\right)^n \qquad 0 \leqslant v \leqslant \theta$

$f_{\bar{V}}(v) = \frac{d}{dv} F_{\bar{V}}(v) = \frac{1}{\theta}\, n \left(\frac{v}{\theta}\right)^{n-1} \quad , 0 \leqslant v \leqslant \theta$

$F(\bar{V}) = \int_0^\theta \frac{n}{\theta} v \left(\frac{v}{\theta}\right)^{n-1} dv = \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \theta \neq \theta \Rightarrow$ تخمینگر MLE biasدار است

$\boxed{\hat{\theta} = \frac{n+1}{n} \max x_i}$

$L = \prod \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \Rightarrow \ell \propto -n\lambda + \sum x_i \log \lambda$ (۳

$\max \ell \Rightarrow -n + \frac{\sum x_i}{\lambda} = 0 \Rightarrow \boxed{\lambda^* = \frac{\sum x_i}{n}}$

4)

(a) In this problem, the parameter needed to be estimated is $\boldsymbol{\mu}$. Given the training data, we have

$$l(\boldsymbol{\mu})p(\boldsymbol{\mu}) = \ln[p(\mathcal{D}|\boldsymbol{\mu})p(\boldsymbol{\mu})]$$

where for the Gaussian

$$\begin{aligned} \ln[p(\mathcal{D}|\boldsymbol{\mu})] &= \ln\left(\prod_{k=1}^n p(\mathbf{x}_k|\boldsymbol{\mu})\right) \\ &= \sum_{k=1}^n \ln[p(\mathbf{x}_k|\boldsymbol{\mu})] \\ &= -\frac{n}{2}\ln\left[(2\pi)^d|\boldsymbol{\Sigma}|\right] - \sum_{k=1}^n \frac{1}{2}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k - \boldsymbol{\mu}) \end{aligned}$$

and

$$p(\boldsymbol{\mu}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_0|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\mu} - \mathbf{m}_0)^t \boldsymbol{\Sigma}_o^{-1}(\boldsymbol{\mu} - \mathbf{m}_0)\right].$$

The MAP estimator for the mean is then

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \arg\max_{\boldsymbol{\mu}} \left\{ \left[-\frac{n}{2}\ln\left[(2\pi)^d|\boldsymbol{\Sigma}|\right] - \sum_{k=1}^n \frac{1}{2}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k - \boldsymbol{\mu}) \right] \right. \\ &\quad \left. \times \left[\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_0|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\mu} - \mathbf{m}_0)^t \boldsymbol{\Sigma}_o^{-1}(\boldsymbol{\mu} - \mathbf{m}_0)\right] \right] \right\}. \end{aligned}$$

(b) After the linear transform governed by the matrix $\mathbf{A}$, we have

$$\boldsymbol{\mu}' = \mathcal{E}[\mathbf{x}'] = \mathcal{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathcal{E}[\mathbf{x}] = \mathbf{A}\boldsymbol{\mu},$$

and

$$\begin{aligned}
\boldsymbol{\Sigma}' &= \mathcal{E}[(\mathbf{x}'-\boldsymbol{\mu}')(\mathbf{x}'-\boldsymbol{\mu}')^t] \\
&= \mathcal{E}[(\mathbf{A}\mathbf{x}'-\mathbf{A}\boldsymbol{\mu}')(\mathbf{A}\mathbf{x}'-\mathbf{A}\boldsymbol{\mu}')^t] \\
&= \mathcal{E}[\mathbf{A}(\mathbf{x}'-\boldsymbol{\mu}')(\mathbf{x}'-\boldsymbol{\mu}')^t\mathbf{A}^t] \\
&= \mathbf{A}\mathcal{E}[(\mathbf{x}'-\boldsymbol{\mu}')(\mathbf{x}'-\boldsymbol{\mu}')^t]\mathbf{A}^t \\
&= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t.
\end{aligned}$$

Thus we have the log-likelihood

$$\begin{aligned}
\ln[p(\mathcal{D}'|\boldsymbol{\mu}')] &= \ln\left(\prod_{k=1}^{n} p(\mathbf{x}'_k|\boldsymbol{\mu}')\right) \\
&= \ln\left(\prod_{k=1}^{n} p(\mathbf{A}\mathbf{x}_k|\mathbf{A}\boldsymbol{\mu})\right) \\
&= \sum_{k=1}^{n} \ln[p(\mathbf{A}\mathbf{x}_k|\mathbf{A}\boldsymbol{\mu})] \\
&= -\frac{n}{2}\ln\left[(2\pi)^d|\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t|\right] - \sum_{k=1}^{n}\frac{1}{2}(\mathbf{A}\mathbf{x}_k-\mathbf{A}\boldsymbol{\mu})^t(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t)^{-1}(\mathbf{A}\mathbf{x}_k-\mathbf{A}\boldsymbol{\mu}) \\
&= -\frac{n}{2}\ln\left[(2\pi)^d|\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t|\right] - \sum_{k=1}^{n}\frac{1}{2}((\mathbf{x}-\boldsymbol{\mu})^t\mathbf{A}^t)((\mathbf{A}^{-1})^t\boldsymbol{\Sigma}^{-1}\mathbf{A}^{-1})(\mathbf{A}(\mathbf{x}_k-\boldsymbol{\mu})) \\
&= -\frac{n}{2}\ln\left[(2\pi)^d|\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t|\right] - \sum_{k=1}^{n}\frac{1}{2}(\mathbf{x}_k-\boldsymbol{\mu})^t(\mathbf{A}^t(\mathbf{A}^{-1})^t)\boldsymbol{\Sigma}^{-1}(\mathbf{A}^{-1}\mathbf{A})(\mathbf{x}_k-\boldsymbol{\mu}) \\
&= -\frac{n}{2}\ln\left[(2\pi)^d|\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t|\right] - \sum_{k=1}^{n}\frac{1}{2}(\mathbf{x}_k-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}_k-\boldsymbol{\mu}).
\end{aligned}$$

Likewise we have that the density of $\boldsymbol{\mu}'$ is a Gaussian of the form

$$\begin{aligned}
p(\boldsymbol{\mu}') &= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}'_0|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}'-\mathbf{m}_0)^t\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}'-\mathbf{m}_0)\right] \\
&= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}'_0|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{A}\boldsymbol{\mu}-\mathbf{A}\mathbf{m}_0)^t(\mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}^t)^{-1}(\mathbf{A}\boldsymbol{\mu}-\mathbf{A}\mathbf{m}_0)\right] \\
&= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}'_0|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\mathbf{m}_0)^t\mathbf{A}^t(\mathbf{A}^{-1})^t\boldsymbol{\Sigma}_0^{-1}\mathbf{A}^{-1}\mathbf{A}(\boldsymbol{\mu}-\mathbf{m}_0)\right] \\
&= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}'_0|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\mathbf{m}_0)^t\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}-\mathbf{m}_0)\right].
\end{aligned}$$

Thus the new MAP estimator is

$$\begin{aligned}
\hat{\boldsymbol{\mu}}' &= \arg\max_{\boldsymbol{\mu}}\Bigg\{-\frac{n}{2}\ln\left[(2\pi)^d|\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t|\right] \\
&\quad -\sum_{k=1}^{n}\frac{1}{2}(\mathbf{x}_k-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}_k-\boldsymbol{\mu})\left[\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}'_0|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\mathbf{m}_0)^t\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}-\mathbf{m}_0)\right]\right]\Bigg\}.
\end{aligned}$$

We compare $\hat{\boldsymbol{\mu}}$ and see that the two equations are the same, up to a constant. Therefore the estimator gives the appropriate estimate for the transformed mean $\hat{\boldsymbol{\mu}}'$.