2018 Section 1, Problem 2

Amw

2 a) Derive the Bayes rule for classifying a new obs. $x \in \mathbb{R}^p$.

Under 0-1 loss, the Bayes rule is the posterior mode.

The Bayes rule assigns x to $\theta = 0 \iff f(\theta=0|x) > f(\theta=1|x)$

x to $\theta = 1 \iff f(\theta=1|x) > f(\theta=0|x)$

Know $f(\theta=0|x) \propto f(x|\theta=0) \cdot \lambda(\theta=0)$

$\propto \frac{1}{2} \exp\{-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1}(x-\mu_0)\}$

$f(\theta=1|x) \propto f(x|\theta=1) \cdot \lambda(\theta=1)$

$\propto \frac{1}{2} \exp\{-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)\}$

Thus, the Bayes rule assigns x to $\theta=0 \iff \frac{1}{2}\exp\{-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\} > \frac{1}{2}\exp\{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\}$

x to $\theta = 1 \iff \frac{1}{2}\exp\{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\} > \frac{1}{2}\exp\{-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\}$

Look at first case → same work will apply to 2nd.

$-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) > -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$

$\Rightarrow (x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - (x-\mu_0)^T\Sigma^{-1}(x-\mu_0) > 0$

$\Rightarrow \cancel{x^T\Sigma^{-1}x} - 2\mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1 - \cancel{x^T\Sigma^{-1}x} + 2\mu_0^T\Sigma^{-1}x - \mu_0^T\Sigma^{-1}\mu_0 > 0$

$\Rightarrow \mu_1^T\Sigma^{-1}\mu_1 - \mu_0^T\Sigma^{-1}\mu_0 - 2x^T\Sigma^{-1}(\mu_1-\mu_0) > 0$

$\Rightarrow (\mu_1+\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0) - 2x^T\Sigma^{-1}(\mu_1-\mu_0) > 0$

$\Rightarrow [(\mu_1+\mu_0)^T - 2x^T]\Sigma^{-1}(\mu_1-\mu_0) > 0$

$\Rightarrow (\mu_1-\mu_0)^T\Sigma^{-1}[x - (\mu_1+\mu_0)/2] > 0$

$\Rightarrow \delta^T\Sigma^{-1}(x-\mu) > 0$ for $\delta = \mu_1 - \mu_0$ and $\mu = \frac{\mu_1+\mu_0}{2}$

Thus, the $\boxed{\text{Bayes rule} = \begin{cases} \theta = 0 & \text{if } \delta^T\Sigma^{-1}(x-\mu) > 0 \\ \theta = 1 & \text{if } \delta^T\Sigma^{-1}(x-\mu) \le 0 \end{cases}}$

A MW

2018, section 1, Problem 2

2 b) Derive the misclassification rate R^* of the Bayes rule.

⌈ Misclassification rate of Bayes rule = $R^* = P(\text{classify wrong})$

$= P(\text{choice}=1)\cdot P(\text{classify wrong} \mid \text{choice}=1) + P(\text{choice}=0)\cdot P(\text{classify wrong} \mid \text{choice}=0)$

$= P(\theta=1)\cdot P(\delta^T\Sigma^{-1}(x-\mu) > 0 \mid \theta=1) + P(\theta=0)\cdot P(\delta^T\Sigma^{-1}(x-\mu) \leq 0 \mid \theta=0)$

$$= \frac{1}{2} P\left(\underbrace{\frac{\delta^T\Sigma^{-1}(x-\mu) - \delta^T\Sigma^{-1}(\mu_1-\mu)}{\sqrt{\delta^T\Sigma^{-1}\delta}}}_{\sim N(0,1)} > \frac{-\delta^T\Sigma^{-1}(\mu_1-\mu)}{\sqrt{\delta^T\Sigma^{-1}\delta}} \right)$$

$$+ \frac{1}{2} P\left(\underbrace{\frac{\delta^T\Sigma^{-1}(x-\mu) - \delta^T\Sigma^{-1}(\mu_0-\mu)}{\sqrt{\delta^T\Sigma^{-1}\delta}}}_{\sim N(0,1)} \leq \frac{-\delta^T\Sigma^{-1}(\mu_0-\mu)}{\sqrt{\delta^T\Sigma^{-1}\delta}} \right)$$

$$\mu_1 - \frac{(\mu_1+\mu_0)}{2} = \frac{\mu_1-\mu_0}{2} = -\delta/2 \qquad \mu_0 - \frac{(\mu_1+\mu_0)}{2} = \frac{\mu_0-\mu_1}{2} = \delta/2$$

$$= \frac{1}{2} P\left(z > \frac{-\delta^T\Sigma^{-1}(\mu_1-\mu)}{\sqrt{\delta^T\Sigma^{-1}\delta}}\right) + \frac{1}{2} P\left(z \leq \frac{-\delta^T\Sigma^{-1}(\mu_0-\mu)}{\sqrt{\delta^T\Sigma^{-1}\delta}}\right)$$

$$= \frac{1}{2} P\left(z > \frac{+\delta^T\Sigma^{-1}\delta}{2\sqrt{\delta^T\Sigma^{-1}\delta}}\right) + \frac{1}{2} P\left(z \leq \frac{-\delta^T\Sigma^{-1}\delta}{2\sqrt{\delta^T\Sigma^{-1}\delta}}\right)$$

$$= P\left(z \leq \frac{-\delta^T\Sigma^{-1}\delta}{2\sqrt{\delta^T\Sigma^{-1}\delta}}\right) = \Phi\left(-\frac{1}{2}\sqrt{\delta^T\Sigma^{-1}\delta}\right)$$

$\uparrow$ cdf of $N(0,1)$. ⌋

2018 Section 1, Problem 2

2c) Let X_{0i} $(i=1,\dots,n_0)$ be iid samples from the class $\theta=0$ and X_{1i} $(i=1,\dots,n_1)$ be iid samples from the class of $\theta=1$, and X_{0i} is independent of X_{1i}.

Derive the MLEs $(\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma})$ of (μ_0, μ_1, Σ)

$$\mathcal{L}(\mu_0,\mu_1,\Sigma \mid x_0, x_1) = \prod_{i=1}^{n_0} (2\pi|\Sigma|)^{-1/2} \exp\left\{-\tfrac{1}{2}(x_{0i}-\mu_0)^T \Sigma^{-1}(x_{0i}-\mu_0)\right\}$$

$$\cdot \prod_{i=1}^{n_1} (2\pi|\Sigma|)^{-1/2} \exp\left\{-\tfrac{1}{2}(x_{1i}-\mu_1)^T \Sigma^{-1}(x_{1i}-\mu_1)\right\}$$

$$\propto \Sigma^{-n} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n_0}(x_{0i}-\mu_0)^T\Sigma^{-1}(x_{0i}-\mu_0)\right\} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n_0}(x_{1i}-\mu_1)^T\Sigma^{-1}(x_{1i}-\mu_1)\right\}$$

$$\Rightarrow \ell(\mu_0,\mu_1,\Sigma \mid x_0,x_1) = -n\log(|\Sigma|) - \tfrac{1}{2}\sum_{i=1}^{n_0}(x_{0i}-\mu_0)^T\Sigma^{-1}(x_{0i}-\mu_0) - \tfrac{1}{2}\sum_{i=1}^{n_1}(x_{1i}-\mu_1)^T\Sigma^{-1}(x_{1i}-\mu_1)$$

Then, $\dfrac{\partial \ell}{\partial \mu_0} = +\sum_{i=1}^{n_0}\Sigma^{-1}(x_{0i}-\mu_0) \overset{\text{set}}{=} 0 \Rightarrow \sum_{i=1}^{n_0} x_{0i} - \mu_0 n_0 = 0$

$$\Rightarrow \boxed{\hat{\mu}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0} x_{0i}}$$

(Note: $\dfrac{\partial^2 \ell}{\partial \mu_0^2} = -n_0\Sigma^{-1} < 0$ $\Rightarrow \hat{\mu}_0$ occurs at a local max.)

Similarly, $\boxed{\hat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}}$

To find MLE of Σ, first rewrite log likelihood as:

$$\ell(\mu_0,\mu_1,\Sigma \mid x_0,x_1) = \left(\tfrac{n_0}{2}+\tfrac{n_1}{2}\right)\log|\Sigma^{-1}| - \tfrac{1}{2}\sum_{i=1}^{n_0} tr\left[(x_{0i}-\mu_0)(x_{0i}-\mu_0)^T\Sigma^{-1}\right]$$

$$- \tfrac{1}{2}\sum_{i=1}^{n_1} tr\left[(x_{1i}-\mu_1)(x_{1i}-\mu_1)^T\Sigma^{-1}\right]$$

b/c $\dfrac{\partial \log|M|}{\partial M} = \dfrac{1}{M}$

$$\Rightarrow \frac{\partial \ell}{\partial \Sigma^{-1}} = \left(\tfrac{n_0}{2}+\tfrac{n_1}{2}\right)\Sigma - \tfrac{1}{2}\sum_{i=1}^{n_0}(x_{0i}-\mu_0)(x_{0i}-\mu_0)' - \tfrac{1}{2}\sum_{i=1}^{n_1}(x_{1i}-\mu_1)(x_{1i}-\mu_1)' \overset{\text{set}}{=} 0$$

since $\dfrac{1}{\Sigma^{-1}} = \Sigma$

Note derivative w.r.t. Σ^{-1}

$$\Rightarrow \hat{\Sigma} = \frac{1}{(n_0+n_1)}\left[\sum_{i=1}^{n_0}(x_{0i}-\hat{\mu}_0)(x_{0i}-\hat{\mu}_0)' + \sum_{i=1}^{n_1}(x_{1i}-\hat{\mu}_1)(x_{1i}-\hat{\mu}_1)'\right]$$

$$\Rightarrow \boxed{\hat{\Sigma} = \frac{1}{(n_0+n_1)}\left[\sum_{i=1}^{n_0}(x_{0i}-\bar{x}_0)(x_{0i}-\bar{x}_0)' + \sum_{i=1}^{n_1}(x_{1i}-\bar{x}_1)(x_{1i}-\bar{x}_1)'\right]}$$

AMW

2018 Section 1, Problem 2

2.d) If we replace (μ_0, μ_1, Σ) in the Bayes rule w/ $(\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma})$, prove that the misclassification rate of the resulting rule, i.e., the probability of classifying x to a wrong class given the training data $\{x_{0i}\}_{i=1}^{n_0}$ and $\{x_{1i}\}_{i=1}^{n_1}$ is given by:

$$\frac{1}{2}\Phi\left(\frac{\hat{\delta}^T\hat{\Sigma}^{-1}(\mu_1-\hat{\mu})}{\sqrt{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right) + \frac{1}{2}\Phi\left(-\frac{\hat{\delta}^T\hat{\Sigma}^{-1}(\mu_0-\hat{\mu})}{\sqrt{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)$$

where $\hat{\delta} = \hat{\mu}_0 - \hat{\mu}_1$ and $\hat{\mu} = (\hat{\mu}_0 + \hat{\mu}_1)/2$

⌈ Know from $R^* = \frac{1}{2}P\left(z > \frac{-\delta^T\Sigma^{-1}(\mu_1-\mu)}{\sqrt{\delta^T\Sigma^{-1}\Sigma\Sigma^{-1}\delta}}\right) + \frac{1}{2}P\left(z \le \frac{-\delta^T\Sigma^{-1}(\mu_0-\mu)}{\sqrt{\delta^T\Sigma^{-1}\Sigma\Sigma^{-1}\delta}}\right)$

$$= \frac{1}{2}P\left(-z < \frac{+\delta^T\Sigma^{-1}(\mu_1-\mu)}{\sqrt{\delta^T\Sigma^{-1}\Sigma\Sigma^{-1}\delta}}\right) + \frac{1}{2}P\left(z \le \frac{-\delta^T\Sigma^{-1}(\mu_0-\mu)}{\sqrt{\delta^T\Sigma^{-1}\Sigma\Sigma^{-1}\delta}}\right)$$

Then, if we replace (μ_0, μ_1, Σ) w/ $(\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma})$ get:

$$R_n = \frac{1}{2}\Phi\left(\frac{\hat{\delta}^T\hat{\Sigma}^{-1}(\mu_1-\hat{\mu})}{\sqrt{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right) + \frac{1}{2}\Phi\left(\frac{-\hat{\delta}^T\hat{\Sigma}^{-1}(\mu_0-\hat{\mu})}{\sqrt{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)$$

where $\hat{\delta} = \hat{\mu}_0 - \hat{\mu}_1$ and $\hat{\mu} = (\hat{\mu}_0 + \hat{\mu}_1)/2$ as described in a). ⌋

Cont'd. ⟶

AMW

2018, Section 1, Problem 2

2e) We propose another classification rule that assigns X to the class of $\theta=0$ iff $\hat{\beta}^T(x-\hat{\mu}) \geq 0$ where $\hat{\mu} = (\hat{\mu}_0+\hat{\mu}_1)/2$ and $\hat{\beta}$ solves the following problem

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \ \frac{1}{2}\beta^T\hat{\Sigma}\beta - (\hat{\mu}_0-\hat{\mu}_1)^T\beta + \lambda\sum_{j=1}^{p}|\beta_j|$$

Derive the M-M algorithm for solving $\hat{\beta}$. Give an explicit choice of step size and closed-form expressions on how iterations need to be done.

① Find the majorization function

② Minimize the majorization function together w/ the penalty function

① Objective fcn: $\frac{1}{2}\beta^T\hat{\Sigma}\beta - (\hat{\mu}_0-\hat{\mu}_1)^T\beta + \lambda\sum_{j=1}^{p}|\beta_j|$

Let $l(\beta) = \frac{1}{2}\beta^T\hat{\Sigma}\beta - (\hat{\mu}_0-\hat{\mu}_1)^T\beta$ and $g(\beta) = \lambda\sum_{j=1}^{p}|\beta_j|$

$\Rightarrow \nabla l(\beta) = \hat{\Sigma}\beta - (\hat{\mu}_0-\hat{\mu}_1)$

$\Rightarrow \nabla^2 l(\beta) = \hat{\Sigma}$

Then, a Taylor expansion around $\tilde{\beta}$ gives

$$l(\beta) = l(\tilde{\beta}) + \nabla l(\tilde{\beta})^T(\beta-\tilde{\beta}) + \frac{1}{2}(\beta-\tilde{\beta})^T\nabla^2 l(\tilde{\beta})(\beta-\tilde{\beta})$$

$$\leq \underbrace{l(\tilde{\beta}) + [\hat{\Sigma}\tilde{\beta} - (\hat{\mu}_0-\hat{\mu}_1)]^T(\beta-\tilde{\beta})_j + \frac{c}{2}(\beta-\tilde{\beta})^T(\beta-\tilde{\beta})}_{l_Q(\beta)}$$

where $c = \sup_{\beta} \lambda_{max}(\nabla^2 l(\beta)) = \sup_{\beta} \lambda_{max}(\hat{\Sigma})$

add back penalty term

② Then, $\tilde{\beta}^{(new)} = \underset{\beta}{\operatorname{argmin}}\left[\underbrace{l_Q(\beta)}_{\text{majorization fcn.}} + \lambda\sum_{j=1}^{p}|\beta_j|\right]$

$$\Rightarrow f(\beta_j) = l(\tilde{\beta}_j) + [\hat{\Sigma}\tilde{\beta} - (\hat{\mu}_0-\hat{\mu}_1)]_j(\beta-\tilde{\beta})_j + \frac{c}{2}(\beta-\tilde{\beta})_j^T(\beta-\tilde{\beta})_j + \lambda\frac{\partial|\beta_j|}{\partial\beta_j}$$

$$\Rightarrow \frac{\partial f}{\partial\beta_j} = [\hat{\Sigma}\tilde{\beta} - \underbrace{(\hat{\mu}_0-\hat{\mu}_1)}_{\hat{\delta}}]_j + c(\beta-\tilde{\beta})_j + \lambda\frac{\partial|\beta_j|}{\partial\beta_j} = \begin{cases} -1 & \text{if } \beta_j<0 \\ [-1,1] & \text{if } \beta_j=0 \\ 1 & \text{if } \beta_j>0 \end{cases}$$

Cont'd. $\longrightarrow$

2018, Section 1, Problem 2

2e) cont'd

By KKT conditions, know $0 \in \left.\frac{\partial f}{\partial \beta_j}\right|_{\beta_j = \tilde{\beta}_j^{(new)}}$.

For $\tilde{\beta}_j^{(new)} < 0$: $[\hat{\Sigma}\tilde{\beta} - \hat{\delta}]_j + c\tilde{\beta}_j^{(new)} - c\tilde{\beta}_j + \lambda(-1) \overset{set}{=} 0$

$$\Rightarrow c\tilde{\beta}_j^{(new)} = \lambda + c\tilde{\beta}_j - [\hat{\Sigma}\tilde{\beta} - \hat{\delta}]_j$$

$$\Rightarrow \tilde{\beta}_j^{(new)} = \tilde{\beta}_j - \underbrace{\tfrac{1}{c}}_{\text{step size}}[\hat{\Sigma}\tilde{\beta} - \hat{\delta}]_j + \lambda/c$$

For $\tilde{\beta}_j^{(new)} = 0$: $0 \in (\hat{\Sigma}\tilde{\beta} - \hat{\delta})_j + c(\overset{=0}{\tilde{\beta}_j^{(new)}} - \tilde{\beta})_j + \lambda[-1,1]$

divide by c to match form in last part

$$\Rightarrow 0 \in \tfrac{1}{c}(\hat{\Sigma}\tilde{\beta} - \hat{\delta})_j - \tilde{\beta}_j + [-\lambda/c, \lambda/c]$$

by symmetry of $[-\lambda/c, \lambda/c]$

$$\Rightarrow \tilde{\beta}_j - \tfrac{1}{c}(\hat{\Sigma}\tilde{\beta} - \hat{\delta})_j \in [-\lambda/c, \lambda/c]$$

For $\tilde{\beta}_j^{(new)} > 0$: Similar to first one, but $-\lambda/c$ instead of λ/c:

$$\tilde{\beta}_j^{(new)} = \tilde{\beta}_j - \tfrac{1}{c}[\hat{\Sigma}\tilde{\beta} - \hat{\delta}]_j - \lambda/c$$

Thus, $\boxed{\tilde{\beta}_j^{(new)} = S(\underbrace{\tilde{\beta}_j - \tfrac{1}{c}[\hat{\Sigma}\tilde{\beta} - \hat{\delta}]_j}_{z}, \lambda/c)}$

where $S(z, \lambda/c) = \text{sign}(z)(|z| - \lambda/c)_+$ or $S(z, \lambda/c) = \begin{cases} z - \lambda/c, & z > 0, \ \lambda < |z| \\ z + \lambda/c, & z < 0, \ \lambda < |z| \\ 0, & \lambda \geq |z| \end{cases}$

Algorithm:

1. Initialize β to $\beta^{(0)} \in \mathbb{R}^p$. Set $k=1$
2. Compute $\beta^{(k)} = S(\beta^{(k-1)} - \frac{1}{c}[\hat{\Sigma}\beta^{(k-1)} - \hat{\delta}], \lambda/c)$ for $S(z,\lambda)$ the soft-thresholding fcn described above.
3. Iterate until convergence w/in a prespecified tolerance $\|\beta^{(k)} - \beta^{(k-1)}\|_2 < \epsilon$ (e.g. $\epsilon = 10^{-4}$).

AMW

2018, Section 1, Problem 2

2f) Let R_n denote the misclassification rate of the rule described in e).

Suppose we can show that $\hat{\beta} \xrightarrow{P} \Sigma^{-1}(\mu_0 - \mu_1)$ as $n \to \infty$.

Use this result to prove $R_n \xrightarrow{P} R^*$.

⌈ Know from d) that misclassification rate $= \frac{1}{2}\Phi\left(\frac{\underbrace{\hat{\delta}^T}_{=(\hat{\mu}_0-\hat{\mu}_1)^T}\hat{\Sigma}^{-1}(\mu_1 - \hat{\mu})}{\sqrt{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right) + \frac{1}{2}\Phi\left(-\frac{\underbrace{\hat{\delta}^T}_{=(\hat{\mu}_0-\hat{\mu}_1)^T}\hat{\Sigma}^{-1}(\mu_0 - \hat{\mu})}{\sqrt{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)$

$$= \frac{1}{2}\Phi\left(\frac{(\hat{\mu}_0-\hat{\mu}_1)^T\hat{\Sigma}^{-1}(\mu_1-\hat{\mu})}{\sqrt{(\hat{\mu}_0-\hat{\mu}_1)^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}(\hat{\mu}_0-\hat{\mu}_1)}}\right) + \frac{1}{2}\Phi\left(-\frac{(\hat{\mu}_0-\hat{\mu}_1)^T\hat{\Sigma}^{-1}(\mu_0-\hat{\mu})}{\sqrt{(\hat{\mu}_0-\hat{\mu}_1)^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}(\hat{\mu}_0-\hat{\mu}_1)}}\right)$$

For $\hat{\beta} = (\hat{\mu}_0 - \hat{\mu}_1)^T\hat{\Sigma}^{-1}$

$$= \frac{1}{2}\Phi\left(\frac{\hat{\beta}^T(\mu_1-\hat{\mu})}{\sqrt{\hat{\beta}^T\Sigma\hat{\beta}}}\right) + \frac{1}{2}\Phi\left(\frac{-\hat{\beta}^T(\mu_0-\hat{\mu})}{\sqrt{\hat{\beta}^T\Sigma\hat{\beta}}}\right)$$

By LLN, $\hat{\mu} \xrightarrow{P} \mu$ and $\hat{\beta} \xrightarrow{P} \underbrace{\beta}_{=\Sigma^{-1}(\mu_0-\mu_1)} \Rightarrow R_n \xrightarrow{P} R^*$ as $n \to \infty$.

where

$$R^* = \frac{1}{2}\Phi\left(\frac{(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_1-\mu)}{\sqrt{(\mu_0-\mu_1)^T\underbrace{\Sigma^{-1}\Sigma\Sigma^{-1}}_{\Sigma^{-1}}(\mu_0-\mu_1)}}\right) + \frac{1}{2}\Phi\left(\frac{(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu)}{\sqrt{(\mu_0-\mu_1)^T\underbrace{\Sigma^{-1}\Sigma\Sigma^{-1}}_{\Sigma^{-1}}(\mu_0-\mu_1)}}\right)$$ ⌋