

ECE 131A Project Report

Samuel Li

1. Data Imputation

(a) we start by simplifying $\frac{d}{da_i} E_{MMSE}$:

$$\begin{aligned}\frac{d}{da_i} E_{MMSE} &= \frac{d}{da_i} \left[\sum_{i \in K_{miss}} (X_i - a_i)^2 \right] \\ \frac{d}{da_i} E_{MMSE} &= E \left[\sum_{i \in K_{miss}} \frac{d}{da_i} (X_i - a_i)^2 \right] \\ \frac{d}{da_i} E_{MMSE} &= \sum_{i \in K_{miss}} E [2(a_i - x_i)] \\ \frac{d}{da_i} E_{MMSE} &= 2 \sum_{i \in K_{miss}} (E[a_i] - E[x_i])\end{aligned}$$

Because a_i is a constant, we have $E[a_i] = a_i$, $\frac{d}{da_i} E_{MMSE} = 0$ and $E[x_i] = \mu$, we have

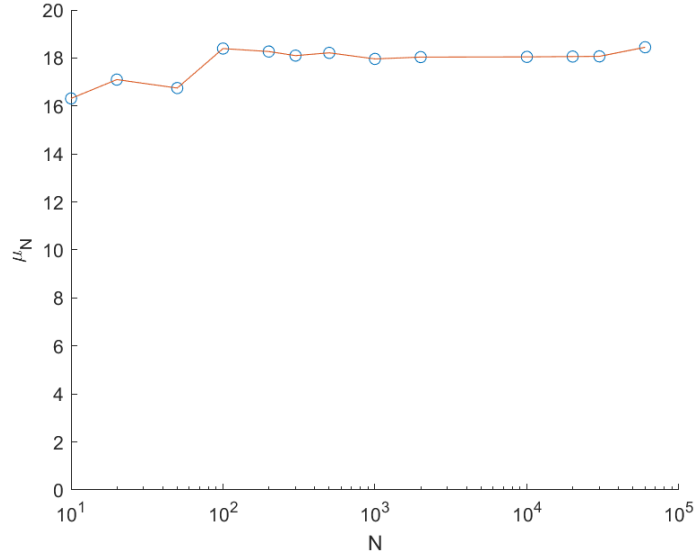
$$2 \sum_{i \in K_{miss}} (a_i - \mu) = 0$$

therefore,

$$a_i = \mu$$

(b)

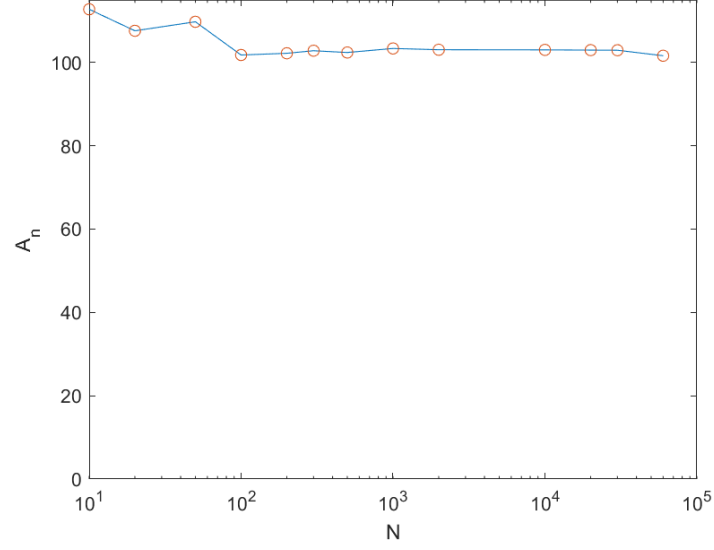
N	10	20	50	100	200	300	500
$\hat{\mu}_N$	16.314	17.099	16.747	18.392	18.269	18.099	18.214
N	1000	2000	10000	20000	30000	60000	-
$\hat{\mu}_N$	17.961	18.037	18.046	18.060	18.067	18.447	-



The empirical estimate of $\hat{\mu}_N$ is consistent across all N's, as the average varies between 16-19. Therefore, we can estimate the average of all N to be around 18.

(c)

N	10	20	50	100	200	300	500
\hat{A}_N	112.721	107.561	109.72	101.753	102.164	102.779	102.356
N	1000	2000	10000	20000	30000	60000	-
\hat{A}_N	103.318	103.016	102.980	102.928	102.898	101.580	-



(d) As N increases, \hat{A}_N approaches around 100. \hat{A}_N does not approaches because as N becomes large, $\hat{A}_N = \sum_{i \in K_{avail}} (x_i - \hat{\mu}_N)^2 / |K_{avail}|$ approaches the variance of x_i which is not necessary 0.

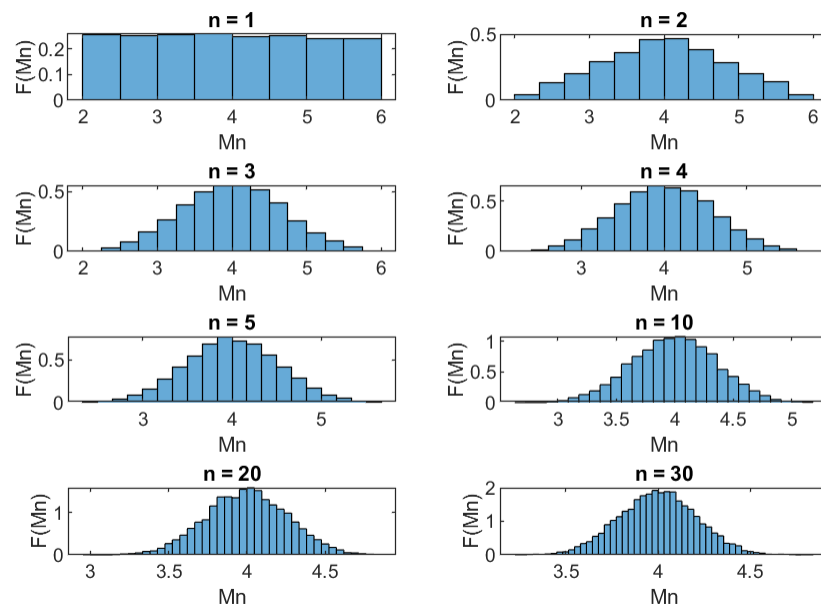
(e) As N becomes large, $\hat{A}_N \rightarrow 100$. Therefore, we can estimate $\sigma^2 \approx 100$.

2. Central Limit Theorem

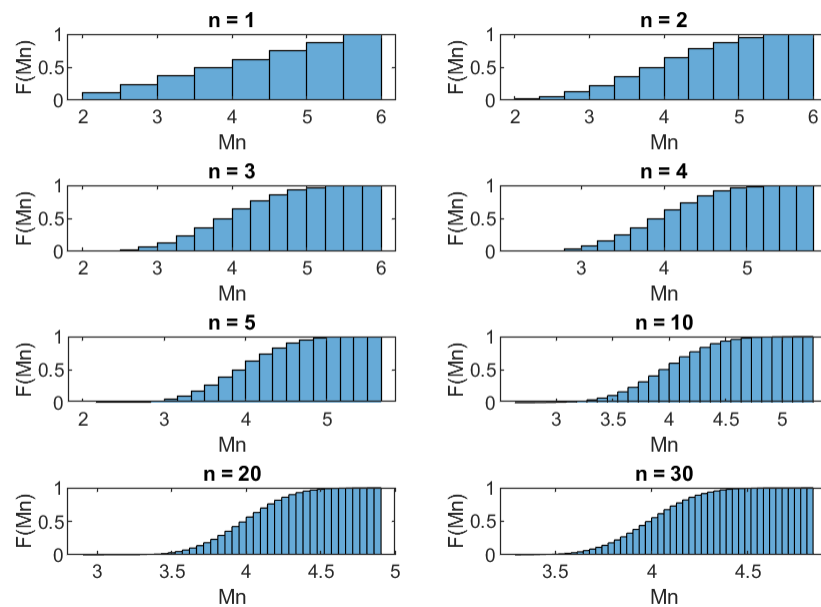
In this Problem, the blue histogram represents data from generated RVs and orange histogram represents fitted gaussian RV.

(a)

PDF for continuous RV



CDF for continuous RV



(b) Because X_i 's are i.i.d uniform RVs, we have

$$\mu = E[X_i] = \frac{2+6}{2} = 4$$

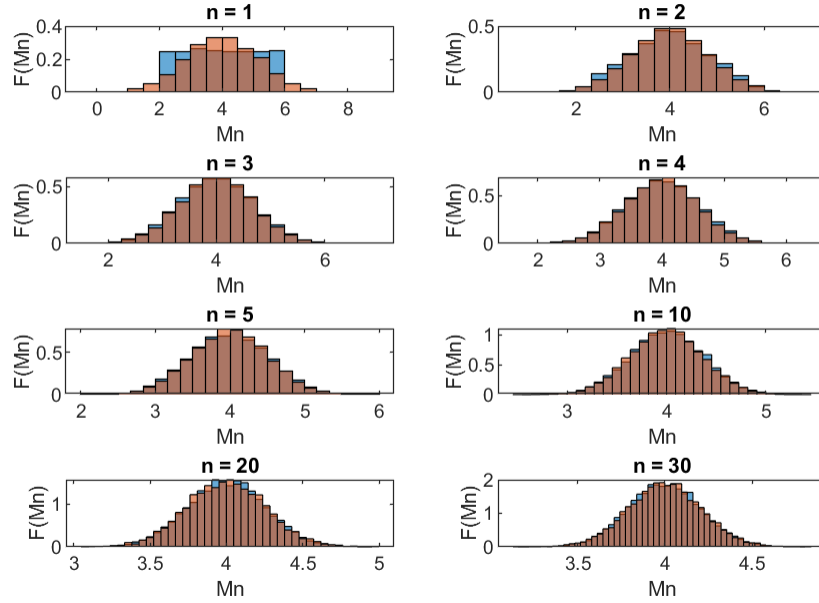
$$\sigma^2 = VAR(X_i) = \frac{(6-2)^2}{12} = \frac{4}{3}$$

$$E[M_n] = \frac{nE[X_i]}{n} = 4$$

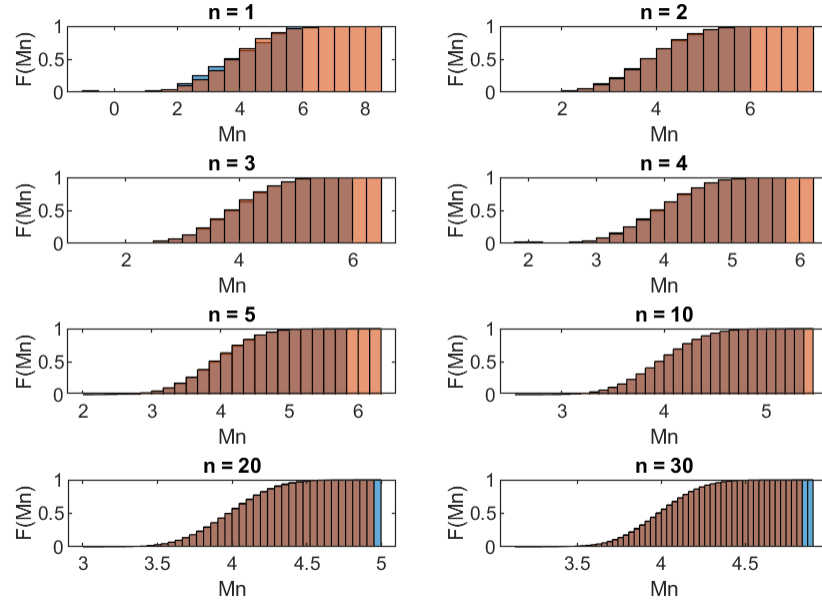
$$VAR(M_n) = VAR\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{4}{3n}$$

(c)

PDF for continuous RV superimposed with gaussian RV



CDF for continuous RV superimposed with gaussian RV

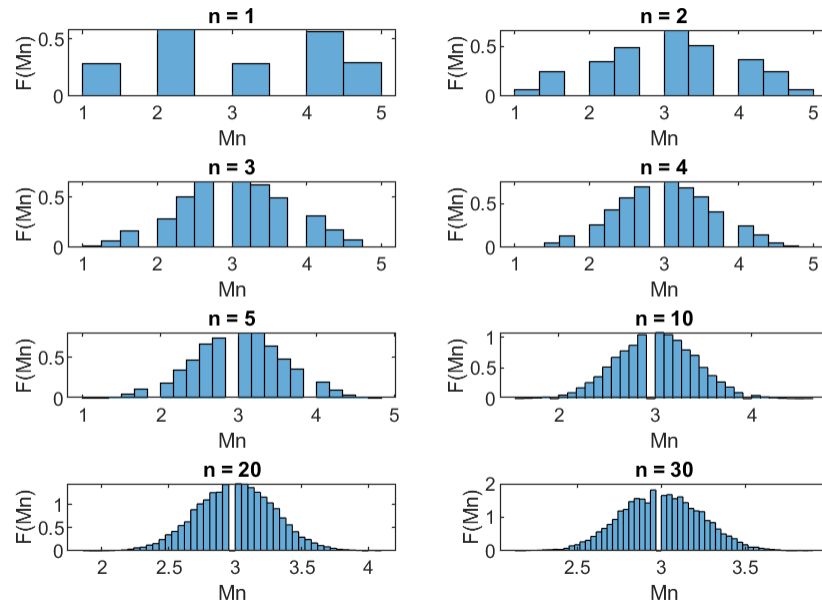


(d)

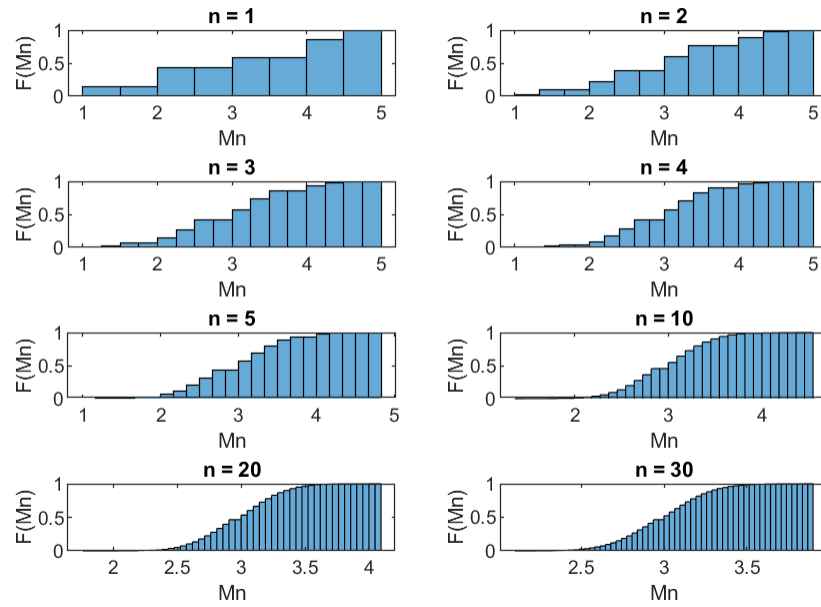
$$E[X_i] = 3$$

$$VAR(X_i) = \frac{12}{7n}$$

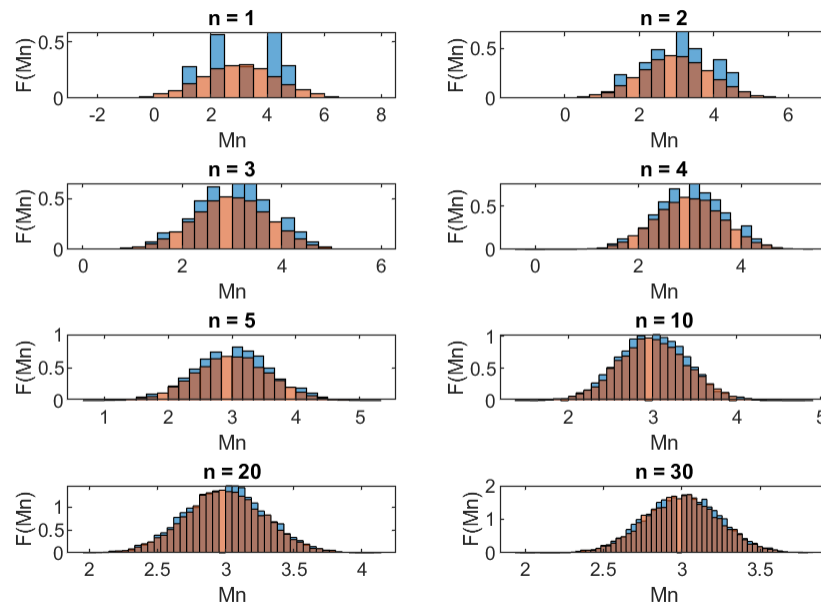
PDF for discrete RV



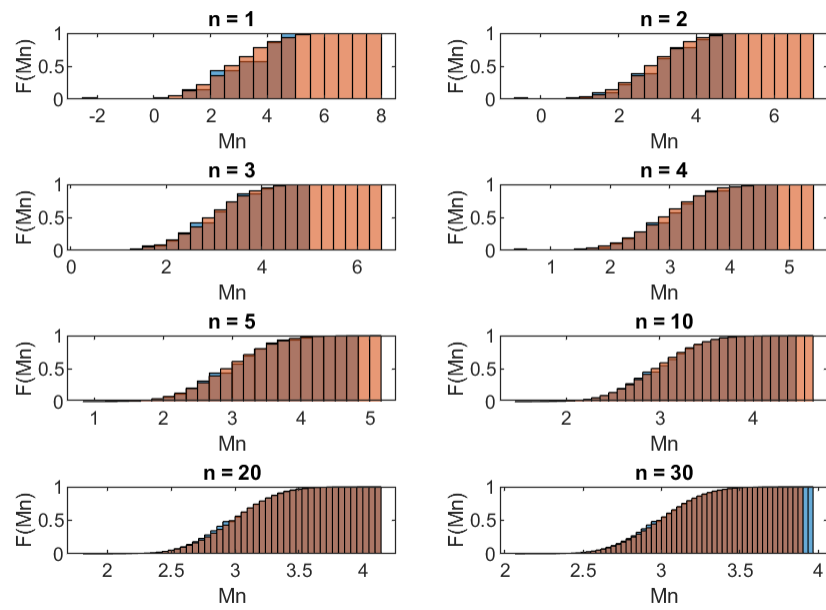
CDF for discrete RV



PDF for discrete RV superimposed with gaussian RV



CDF for discrete RV superimposed with gaussian RV



3. GDA: Gaussian Discriminant Analysis

(a) Using Bayes' Rule, we have

$$P(y = 0|\vec{x}) = \frac{P(\vec{x}|y = 0)P(y = 0)}{P(\vec{x})} = \frac{P(\vec{x}|y = 0)}{2P(\vec{x})}$$

$$P(y = 1|\vec{x}) = \frac{P(\vec{x}|y = 1)P(y = 1)}{P(\vec{x})} = \frac{P(\vec{x}|y = 1)}{2P(\vec{x})}$$

Therefore, we can simplify the inequality as

$$P(\vec{x}|y = 0) \geq P(\vec{x}|y = 1) \quad (3.1)$$

Because

$$f_x = \frac{1}{(2\pi)^{\frac{1}{2}}|\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2}(x - \mu)^{\top} \Sigma^{-1}(x - \mu) \right] \quad (3.2)$$

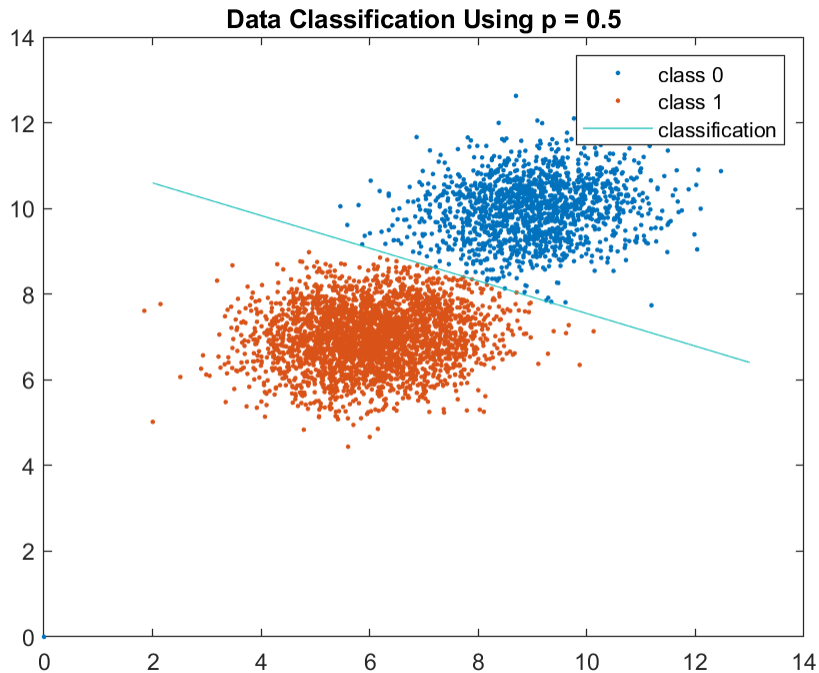
we get

$$\begin{aligned} \frac{1}{(2\pi)^{\frac{1}{2}}|\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2}(x - \mu_0)^{\top} \Sigma^{-1}(x - \mu_0) \right] &\geq \frac{1}{(2\pi)^{\frac{1}{2}}|\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2}(x - \mu_1)^{\top} \Sigma^{-1}(x - \mu_1) \right] \\ -\frac{1}{2}(x - \mu_0)^{\top} \Sigma^{-1}(x - \mu_0) &\geq -\frac{1}{2}(x - \mu_1)^{\top} \Sigma^{-1}(x - \mu_1) \\ (x - \mu_0)^{\top} \Sigma^{-1}(x - \mu_0) &\leq (x - \mu_1)^{\top} \Sigma^{-1}(x - \mu_1) \end{aligned}$$

Because $\Sigma_0 = \Sigma_1 = \Sigma$, $\vec{x}^{\top} \Sigma_0^{-1} \vec{x} = \vec{x}^{\top} \Sigma_1^{-1} \vec{x}$. We get

$$((\vec{\mu}_0 - \vec{\mu}_1)^{\top} \Sigma^{-1}) \cdot \vec{x} + \frac{1}{2} (\vec{\mu}_1^{\top} \Sigma^{-1} \vec{\mu}_1 - \vec{\mu}_0^{\top} \Sigma^{-1} \vec{\mu}_0) \geq 0 \quad (3.3)$$

(b) The percentage of samples identified to be from class 0 is 49.83%.



(c) Using Bayes' Rule, we have

$$P(y = 0|\vec{x}) = \frac{P(\vec{x}|y = 0)P(y = 0)}{P(\vec{x})} = \frac{P(\vec{x}|y = 0)(1 - p)}{P(\vec{x})}$$

$$P(y = 1|\vec{x}) = \frac{P(\vec{x}|y = 1)P(y = 1)}{P(\vec{x})} = \frac{P(\vec{x}|y = 1) \cdot p}{P(\vec{x})}$$

Therefore, the inequality simplifies to

$$P(\vec{x}|y = 0)(1 - p) \geq P(\vec{x}|y = 1) \cdot p \quad (3.4)$$

use Eqn. 3.2, we have

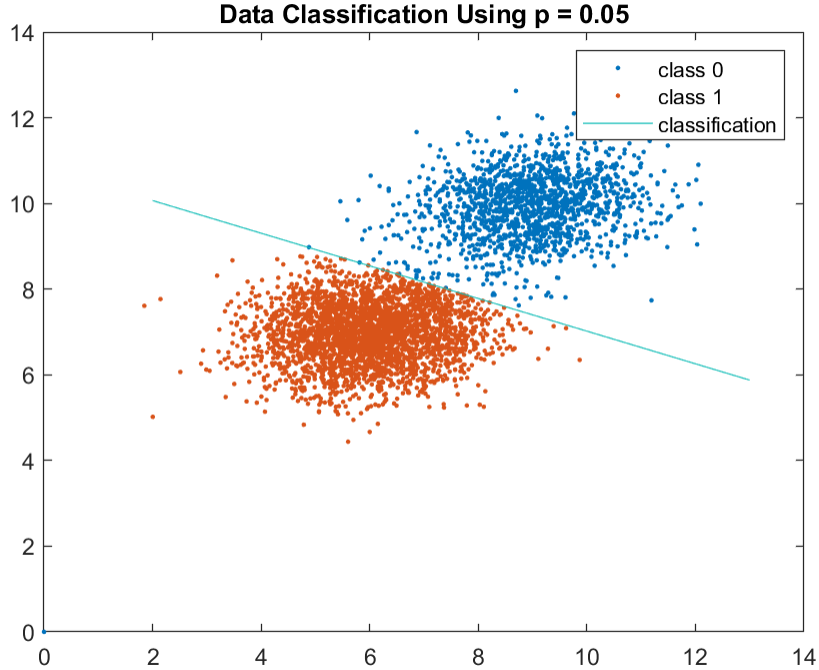
$$\frac{1}{(2\pi)^{|\Sigma|^{\frac{1}{2}}}} \exp \left[-\frac{1}{2}(x - \mu_0)^\top \Sigma^{-1}(x - \mu_0) \right] \cdot p \geq \frac{1}{(2\pi)^{|\Sigma|^{\frac{1}{2}}}} \exp \left[-\frac{1}{2}(x - \mu_1)^\top \Sigma^{-1}(x - \mu_1) \right] (1 - p)$$

$$\ln p + \left[-\frac{1}{2}(x - \mu_0)^\top \Sigma_0^{-1}(x - \mu_0) \right] \geq \ln(1 - p) + \left[-\frac{1}{2}(x - \mu_1)^\top \Sigma_1^{-1}(x - \mu_1) \right]$$

Using $\vec{x}^\top \Sigma_0^{-1} \vec{x} = \vec{x}^\top \Sigma_1^{-1} \vec{x}$,

$$((\vec{\mu}_0 - \vec{\mu}_1) \Sigma^{-1}) \cdot \vec{x} + \frac{1}{2} (\vec{\mu}_1^\top \Sigma^{-1} \vec{\mu}_1 - \vec{\mu}_0^\top \Sigma^{-1} \vec{\mu}_0) + \ln \left(\frac{1 - p}{p} \right) \geq 0 \quad (3.5)$$

(d) The percentage of samples identified to be from class 0 is 51.38%.



The changed p moved the classification line slightly down in the graph.

(e) From Eqn. 3.2 and 3.4, we get

$$\frac{1}{(2\pi)|\Sigma_0|^{\frac{1}{2}}} \exp \left[-\frac{1}{2}(x - \mu_0)^\top \Sigma_0^{-1}(x - \mu_0) \right] \cdot (1 - p) \geq \frac{1}{(2\pi)|\Sigma_1|^{\frac{1}{2}}} \exp \left[-\frac{1}{2}(x - \mu_1)^\top \Sigma_1^{-1}(x - \mu_1) \right] p$$

$$\ln \left(\frac{1-p}{|\Sigma_0|^{\frac{1}{2}}} \right) + \left[-\frac{1}{2}(x - \mu_0)^\top \Sigma_0^{-1}(x - \mu_0) \right] \geq \ln \left(\frac{p}{|\Sigma_1|^{\frac{1}{2}}} \right) + \left[-\frac{1}{2}(x - \mu_1)^\top \Sigma_1^{-1}(x - \mu_1) \right]$$

Expand the equation, we get

$$-\frac{1}{2} \ln \left(\frac{|\Sigma_0|^{\frac{1}{2}}}{1-p} \right) - \frac{1}{2} x^\top \Sigma_0^{-1} x - \frac{1}{2} \mu_0^\top \Sigma_0^{-1} \mu_0 + \mu_0^\top \Sigma_0^{-1} x + \ln(1-p)$$

$$= -\frac{1}{2} \ln \left(\frac{|\Sigma_1|^{\frac{1}{2}}}{p} \right) - \frac{1}{2} x^\top \Sigma_1^{-1} x - \frac{1}{2} \mu_1^\top \Sigma_1^{-1} \mu_1 + \mu_1^\top \Sigma_1^{-1} x + \ln(p)$$

which simplifies to

$$\vec{x}^\top \left(\frac{1}{2} (\Sigma_0^{-1} - \Sigma_1^{-1}) \right) \vec{x} + (\mu_1^\top \Sigma_1^{-1} - \mu_0^\top \Sigma_0^{-1}) \cdot \vec{x} + \frac{1}{2} (\mu_0^\top \Sigma_0^{-1} \mu_0 - \mu_1^\top \Sigma_1^{-1} \mu_1) + \ln \left(\frac{(1-p)|\Sigma_1|^{\frac{1}{2}}}{p|\Sigma_0|^{\frac{1}{2}}} \right) \geq 0 \quad (3.6)$$

(f) The percentage of samples identified to be from class 0 is 50.41%

