EE559-Homework#9-Lei Lei

1. **Convergence of Kernel Density Estimation (KDE)**

(a) $\Phi(\underline{u}) = \exp\left(-\frac{1}{2}\|\underline{u}\|^2\right)$

$= \exp\left(-\frac{1}{2}\underline{u}\underline{u}^T\right)$

$\because \|\underline{u}\|^2 \geq 0$

$\therefore \sup \Phi(\underline{u}) = 1 < \infty \Rightarrow$ satisfy condition (1)

(b) $V_n = \frac{b}{\sqrt{n}}$, $b =$ constant

$\therefore \lim_{n\to\infty} V_n = \frac{b}{\infty} = 0 \Rightarrow$ satisfy condition (3)

$nV_n = n \cdot \frac{b}{\sqrt{n}} = \sqrt{n}\, b$

$\therefore \lim_{n\to\infty} nV_n = \infty b = \infty \Rightarrow$ satisfy condition (4)

(c) $\Phi\left(\frac{\underline{x}}{h_n}\right) = \exp\left(-\frac{1}{2}\left\|\frac{\underline{x}}{h_n}\right\|^2\right) = \exp\left(-\frac{1}{2}(h_n)^{-2}\|\underline{x}\|^2\right)$

$\int \frac{1}{V_n}\Phi\left(\frac{\underline{x}}{h_n}\right)d\underline{x} = 1$

$\frac{1}{V_n}\int \Phi\left(\frac{\underline{x}}{h_n}\right)d\underline{x} = 1 \qquad \int \Phi\left(\frac{\underline{x}}{h_n}\right)d\underline{x} = V_n$

$\therefore V_n = \int \Phi\left(\frac{\underline{x}}{h_n}\right)d\underline{x}$

Let $V_n = \frac{b}{\sqrt{n}}$

$\frac{b}{\sqrt{n}} = \int \Phi\left(\frac{\underline{x}}{h_n}\right)d\underline{x} = \int h_n \Phi\left(\frac{\underline{x}}{h_n}\right)d\frac{\underline{x}}{h_n} = h_n \underbrace{\int \Phi\left(\frac{\underline{x}}{h_n}\right)d\frac{\underline{x}}{h_n}}_{=1}$

$\therefore h_n = \frac{b}{\sqrt{n}}$

2. **2-class minimum-error classification based on KDE estimates**

(a) True Bayes minimum error classifier

(a) $P(\underline{x}|S_1)P(S_1) \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} P(\underline{x}|S_2)P(S_2)$

$P(S_1)[a_1 p_1(\underline{x}) + a_2 p_2(\underline{x})] \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} N(\underline{x}|\underline{m}_2, \underline{\underline{\Sigma}}_2)\, \cancel{P(S_2)}$

$a_1 \cdot N(\underline{x}|\underline{m}_1, \underline{\underline{\Sigma}}_1) + a_2 \cdot U_{x_1}(0,2)\, U_{x_2}(-1,1) \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} N(\underline{x}|\underline{m}_2, \underline{\underline{\Sigma}}_2)$

$U_{x_1}(0,2) = \frac{1}{2} \begin{cases} 1 & 0 \le x_1 \le 2 \\ 0 & \text{otherwise} \end{cases}$

$U_{x_2}(-1,1) = \frac{1}{2} \begin{cases} 1 & -1 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$

when $x_1 \in [0,2]$, $x_2 \in [-1,1]$

$a_1 \cdot N(\underline{x}|\underline{m}_1, \underline{\underline{\Sigma}}_1) + \frac{1}{4} a_2 \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} N(\underline{x}|\underline{m}_2, \underline{\underline{\Sigma}}_2)$

$\ln\left[a_1 N(\underline{x}|\underline{m}_1, \underline{\underline{\Sigma}}_1) + \frac{1}{4} a_2\right] \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} \ln N(\underline{x}|\underline{m}_2, \underline{\underline{\Sigma}}_2)$

$\ln a_1 \cdot \ln \frac{1}{4} a_2 + \ln N(\underline{x}|\underline{m}_1, \underline{\underline{\Sigma}}_1) \ln \frac{1}{4} a_2 \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} \ln N(\underline{x}|\underline{m}_2, \underline{\underline{\Sigma}}_2)$

$\underbrace{\ln a_1 \ln a_2}_{=0} + \ln a_1 \ln \frac{1}{4} + \ln \frac{1}{4} a_2 \left[-\frac{1}{2}\ln|\underline{\underline{\Sigma}}_1| - \frac{1}{2} d_M^2(\underline{x}, \underline{m}_1)\right] \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} -\frac{1}{2}\ln|\underline{\underline{\Sigma}}_2| - \frac{1}{2} d_M^2(\underline{x}, \underline{m}_2)$

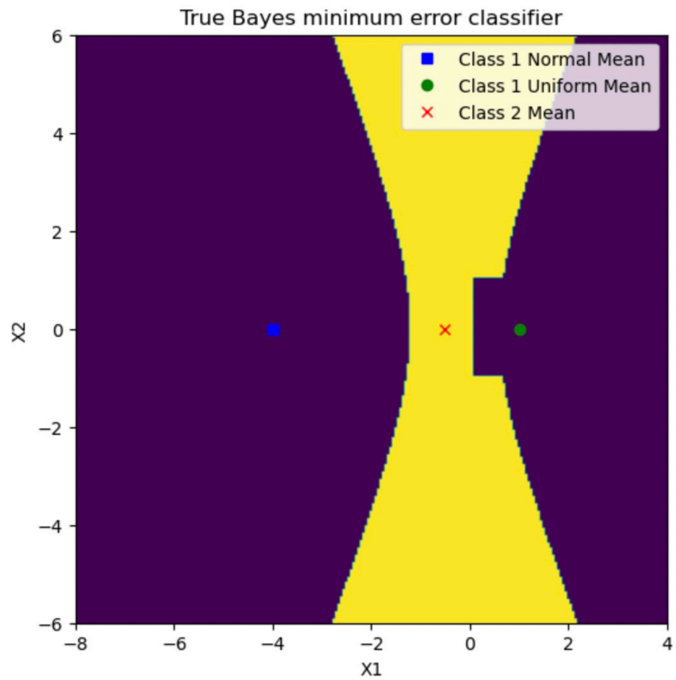
$\ln 0.7 \ln \frac{1}{4} + \ln \frac{0.3}{4} \left[-\frac{1}{2}\ln|\underline{\underline{\Sigma}}_1| - \frac{1}{2} d_M^2(\underline{x}, \underline{m}_1)\right] \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} -\frac{1}{2}\ln|\underline{\underline{\Sigma}}_2| - \frac{1}{2} d_M^2(\underline{x}, \underline{m}_2)$

when otherwise

$a_1 N(\underline{x}|\underline{m}_1, \underline{\underline{\Sigma}}_1) \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} N(\underline{x}|\underline{m}_2, \underline{\underline{\Sigma}}_2)$

$-\frac{1}{2}\ln|\underline{\underline{\Sigma}}_1| - \frac{1}{2} d_M^2(\underline{x}, \underline{m}_1) + \ln a_1 \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} -\frac{1}{2}\ln|\underline{\underline{\Sigma}}_2| - \frac{1}{2} d_M^2(\underline{x}, \underline{m}_2)$

$-\frac{1}{2}\ln|\underline{\underline{\Sigma}}_1| - \frac{1}{2} d_M^2(\underline{x}, \underline{m}_1) + \ln 0.7 \underset{\Gamma_2}{\overset{\Gamma_1}{\gtrless}} -\frac{1}{2}\ln|\underline{\underline{\Sigma}}_2| - \frac{1}{2} d_M^2(\underline{x}, \underline{m}_2)$



(b) Dataset generation.

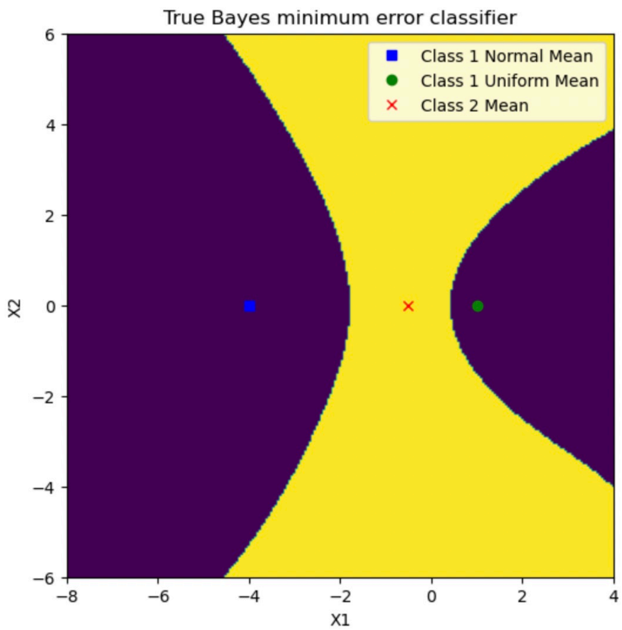
(c) Ideal accuracy.
The accuracy is 0.9601

(d) Minimum-error classifier based on estimates from the training data.
(i)

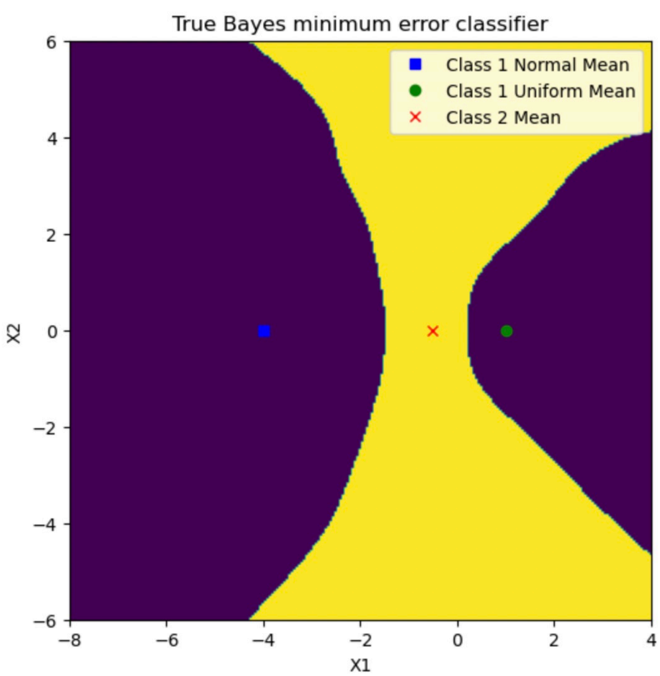
```
When N = 200, P_est(S1) = 0.46, P_est(S2) = 0.54
When N = 2000, P_est(S1) = 0.493, P_est(S2) = 0.507
When N = 20000, P_est(S1) = 0.4935, P_est(S2) = 0.5065
```

(ii)

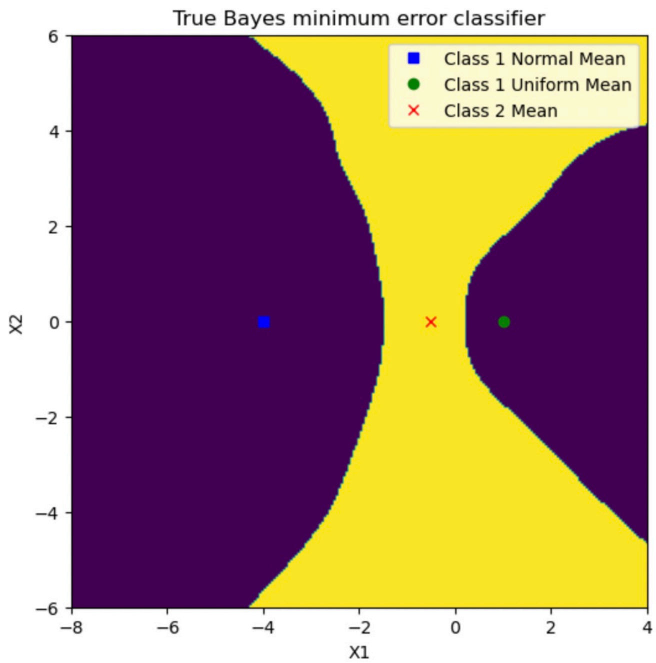
N = 200:



N = 2000:



N = 20000:



(iii)
N = 200: The accuracy is 0.9395
N = 2000: The accuracy is 0.9465
N = 20000: The accuracy is 0.9543

(e) Comparison.
(i) Compare the classification accuracies to each other, and compare the plots to each other. Explain your observations.
Obviously, when N becomes larger, the test accuracy grows. But the largest accuracy didn't reach the one we calculated using the prior P(S1) and P(S2).
And from the plots we can find that while N becomes larger, the resolution decreases because we are using the KDE to get the estimated probability instead of the real one.

(ii) Do the error rates seem consistent with the plots? Explain why or why not.
The error rate seems to be consistent with the plots. While N increasing, the resolution of the classifier decreases, and the accuracy of the classifier increases. Because the decision boundary is approaching the decision boundary we got from the prior in problem (a).