AnalyticalProblems

July 3, 2024

0.1 Question 1

Part I: Analytical problems

1.

In a two-class one-dimensional problem, the pdfs are the Gaussians $\mathcal{N}(0,\sigma^2)$ and $\mathcal{N}(1,\sigma^2)$ for the two classes, respectively. Show that the threshold x_0 minimizing the average risk is equal to

$$x_0 = 1/2 - \sigma^2 \ln \frac{\lambda_{21}P(\omega_2)}{\lambda_{12}P(\omega_1)}$$

where $\lambda_{11} = \lambda_{22} = 0$ has been assumed.

Year:............ Month:............ Day:............

$$1. \ P(x|w_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$P(x|w_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-1)^2}{2\sigma^2}\right) \qquad \lambda = \begin{pmatrix} 0 & \lambda_{12} \\ \lambda_{21} & 0 \end{pmatrix}$$

$$\lambda_{21} \, P(x|w_2) \, P(w_2) = \lambda_{12} \, P(x|w_1) \, P(w_1) \longrightarrow$$

$$\ln \lambda_{21} P(w_2) - \frac{1}{2} \ln 2\pi\sigma^2 - \frac{(x-1)^2}{2\sigma^2} =$$

$$\ln \lambda_{12} P(w_1) - \frac{1}{2} \ln 2\pi\sigma^2 - \frac{x^2}{2\sigma^2} \longrightarrow$$

$$\frac{-x^2 + 2x - 1}{2\sigma^2} + \frac{x^2}{2\sigma^2} = \ln \frac{\lambda_{12} \, P(w_1)}{\lambda_{21} \, P(w_2)} \longrightarrow$$

$$\frac{2x-1}{2} = -\sigma^2 \ln \frac{\lambda_{21} P(w_2)}{\lambda_{12} P(w_1)} \longrightarrow x = \frac{1}{2} - \sigma \ln \frac{\lambda_{21} P(w_2)}{\lambda_{12} P(w_1)}$$

1

0.2 Question 2

2.

Prove that the covariance estimate

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \hat{\mu})(x_k - \hat{\mu})^T$$

is an unbiased one, where

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{N} x_k$$

Year: .... Month: .... Day: ....

2. $\hat{\Sigma} = \frac{1}{N-1} \sum\limits_{k=1}^{N} (x_k - \hat{\mu})(x_k - \hat{\mu})^T \qquad \hat{\mu} = \frac{1}{N} \sum\limits_{k=1}^{N} x_k$

$\forall_{i,j \leq t} \quad \hat{\sigma}_{ij} = \frac{1}{N-1} \sum\limits_{k=1}^{N} (x_k - \hat{\mu}_i)(x_k - \hat{\mu}_j) =$

$\frac{1}{N-1} \sum\limits_{k=1}^{N} x_k^2 - x_k \hat{\mu}_i - x_k \hat{\mu}_j + \hat{\mu}_i \hat{\mu}_j =$

$\frac{1}{N-1} \left(\sum\limits_{k=1}^{N} (x_k)^2 - N\hat{\mu}_i \hat{\mu}_j - N\hat{\mu}_i \hat{\mu}_j + N\hat{\mu}_i \hat{\mu}_j \right) \rightarrow$

$E(\hat{\sigma}_{ij}) = \frac{1}{N-1} \left(\sum\limits_{k=1}^{N} E(x_k^2) - N \cdot E(\hat{\mu}_i \hat{\mu}_j) \right) =$

$\frac{1}{N-1} = N(\sigma_{ij}^2 + \mu^2) - N \left(\frac{\sigma_{ij}}{N} + \mu^2 \right) = \frac{N-1}{N-1} \sigma_{ij}^2 = \sigma_{ij}^2$

note that $E(X^2) = \text{Var}(X) + E(X)^2$

2

0.3 Question 3

3.

Show that for the lognormal distribution

$$p(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \theta)^2}{2\sigma^2}\right), \quad x > 0$$

the ML estimate is given by

$$\hat{\theta}_{ML} = \frac{1}{N} \sum_{k=1}^{N} \ln x_k$$

$$3. \ P(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \theta)^2}{2\sigma^2}\right) \quad x > 0$$

$$P(\bar{x}; \theta) = \prod^{N} P(x) \longrightarrow L(x, \theta) = \sum \ln P(x) =$$

$$\sum_{K=1}^{N} -\ln \sigma - \ln x - \frac{1}{2}\ln 2\pi - \frac{(\ln x_k - \theta)^2}{2\sigma^2} \longrightarrow$$

$$\frac{\partial L}{\partial \theta} = \sum_{K=1}^{N} \frac{2(\ln x_k - \theta)}{2\sigma^2} = 0 \longrightarrow \frac{1}{\sigma^2} \sum_{K=1}^{N} \ln x_k - \theta = 0 \longrightarrow$$

$$\sum_{K=1}^{N} \ln x_k - N\theta = 0 \longrightarrow \hat{\theta} = \frac{1}{N} \sum_{K=1}^{N} \ln x_k$$

3

0.4   Question 4

Consider Cauchy distributions in a two-class one-dimensional classification problem

$$f(x \mid \omega_i) = \frac{1}{\pi b} \cdot \frac{1}{1 + (\frac{x-a_i}{b})^2} \qquad i = 1,2 \qquad a_2 > a_1$$

(a) By explicit integration, check that the distributions are indeed normalized.

(b) Assuming $P(\omega_1) = P(\omega_2)$, show that $P(\omega_1 \mid x) = P(\omega_2 \mid x)$ if $x = (a_1+a_2)/2$. Plot $P(\omega_1 \mid x)$ for the case $a_1 = 3$, $a_2 = 5$ and $b = 1$. How do $P(\omega_1 \mid x)$ and $P(\omega_2 \mid x)$ behave as $x \to -\infty$? $x \to +\infty$? Explain.

(c) Show that the minimum probability of error is given by

$$P(\text{error}) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left| \frac{a_2 - a_1}{2b} \right|$$

Plot this as a function of $|a_2 - a_1|/b$.

(d) What is the maximum value of $P(\text{error})$ and under which conditions can this occur? Explain.

(e) Design the Bayes minimum error classifier in terms of a_i and b if $P(\omega_1) = P(\omega_2)$. Show the decision boundaries in this case. What is the probability of error?

(f) Design the Bayes minimum risk classifier with the following error weights

$$\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$$

Show the decision boundaries in this case. What is the probability of error? Compare the results in (e) and (f).

4. $f(x|w_i) = \dfrac{1}{\pi b} \times \dfrac{1}{1+\left(\frac{x-a_i}{b}\right)^2}$ $a_2 > a_1$

a) $\displaystyle\int_{-\infty}^{+\infty} f(x|w_i)\,dx = \int_{-\infty}^{+\infty} \dfrac{1}{\pi b} \times \dfrac{1}{1+\left(\frac{x-a_i}{b}\right)^2}\,dx$

$\left\{ \dfrac{x-a_i}{b} = x' \rightarrow dx = b\,dx' \right\}$

$= \displaystyle\int_{-\infty}^{+\infty} \dfrac{1}{\pi b} \times \dfrac{1}{1+x'^2}\, b\,dx' = \dfrac{1}{\pi}\int_{-\infty}^{+\infty} \dfrac{1}{1+x'^2}\,dx' = \dfrac{1}{\pi}\left[\arctan(x)\right]_{-\infty}^{+\infty}$

$\arctan(x)\Big]_{-\infty}^{-\infty}\,) = \dfrac{1}{\pi}\left(\dfrac{\pi}{2} - \left(-\dfrac{\pi}{2}\right)\right) = 1$

b) $x = \dfrac{a_1+a_2}{2} \rightarrow P(x|w_1) = \dfrac{1}{\pi b} \times \dfrac{1}{1+\left(\frac{\frac{a_1+a_2}{2}-a_1}{b}\right)^2} =$

$\dfrac{1}{\pi b} \times \dfrac{1}{1+\left(\frac{a_2-a_1}{b}\right)^2}$

$P(x|w_2) = \dfrac{1}{\pi b} \times \dfrac{1}{1+\left(\frac{\frac{a_1+a_2}{2}-a_2}{b}\right)^2} = \dfrac{1}{\pi b} \times \dfrac{1}{1+\left(\frac{a_1-a_2}{b}\right)^2}$

$\rightarrow P(x|w_1) = P(x|w_2)$, $P(w_1) = P(w_2) \rightarrow P(w_2|x) = P(w_1|x)$

$P(w_1|x) = \dfrac{P(x|w_1)P(w_1)}{P(x)} = \dfrac{\frac{1}{2}P(x|w_1)}{\frac{1}{2}P(x|w_1)+\frac{1}{2}P(x|w_2)} =$

$\dfrac{\frac{1}{\pi}\cdot\frac{1}{1+(x-3)^2}}{\frac{1}{\pi}\cdot\frac{1}{1+(x-3)^2}+\frac{1}{\pi}\cdot\frac{1}{1+(x-5)^2}} = \dfrac{x^2-10x+26}{2x^2-16x+36} \rightarrow P(w_2|x) = \dfrac{x^2-6x+10}{2x^2-16x+36}$

$\displaystyle\lim_{x\to+\infty} P(w_1|x) = \dfrac{1}{2}$ $\displaystyle\lim_{x\to-\infty} P(w_1|x) = \dfrac{1}{2}$

$\displaystyle\lim_{x\to+\infty} P(w_2|x) = \dfrac{1}{2}$ $\displaystyle\lim_{x\to-\infty} P(w_2|x) = \dfrac{1}{2}$

c) $x_0 \to P(x_0|w_1) = P(x_0|w_2) \to \dfrac{1}{\pi b} \cdot \dfrac{1}{1 + \left(\frac{x_0 - a_1}{b}\right)^2} = \dfrac{1}{\pi b} \cdot \dfrac{1}{1 + \left(\frac{x_0 - a_2}{b}\right)^2}$

$\to \left(\dfrac{x_0 - a_1}{b}\right)^2 = \left(\dfrac{x_0 - a_2}{b}\right)^2 \to x_0 - a_1 = \pm(x_0 - a_2) \to x_0 = \dfrac{a_1 + a_2}{2}$

$P_e = \displaystyle\int_{-\infty}^{x_0} P(x|w_1) P(w_1)\, dx + \int_{x_0}^{+\infty} P(x|w_2) P(w_2) =$

$P(w_1) \displaystyle\int_{-\infty}^{\frac{a_1 + a_2}{2}} \dfrac{1}{\pi b} \cdot \dfrac{1}{1 + \left(\frac{x - a_1}{b}\right)^2}\, dx + \int_{\frac{a_1 + a_2}{2}}^{+\infty} \dfrac{1}{\pi b} \cdot \dfrac{1}{1 + \left(\frac{x - a_2}{b}\right)^2}\, dx\, P(w_2) =$

$P(w_1)\left[\dfrac{1}{\pi}\arctan\left(\dfrac{a_2 - a_1}{2b}\right) - \dfrac{1}{\pi} \times \left(\dfrac{-\pi}{2}\right)\right] + P(w_2)\left[\dfrac{1}{\pi} \times \dfrac{\pi}{2} - \right.$

$\dfrac{1}{\pi}\arctan\left(\dfrac{a_1 - a_2}{2b}\right)\Big] = P(w_1)\left[\dfrac{1}{2} + \dfrac{1}{\pi}\tan^{-1}\left(\dfrac{a_2 - a_1}{2b}\right)\right] +$

$(1 - P(w_1))\left[\dfrac{1}{2} - \dfrac{1}{\pi}\tan^{-1}\left(\dfrac{a_1 - a_2}{2b}\right)\right] = \dfrac{1}{2} - \dfrac{1}{\pi}\tan^{-1}\left|\dfrac{a_2 - a_1}{2b}\right|$

d) $P(\text{error}) = \dfrac{1}{2} - \dfrac{1}{\pi}\tan^{-1}\left|\dfrac{a_2 - a_1}{2b}\right|$

$a_2 = a_1 \to \min(P(\text{error})) = \dfrac{1}{2} - 0 = \dfrac{1}{2}$

e) $P_e = P(w_2) \displaystyle\int_{-\infty}^{x_0} P(x|w_2)\, dx + P(w_1) \int_{x_0}^{+\infty} P(x|w_1)\, dx$
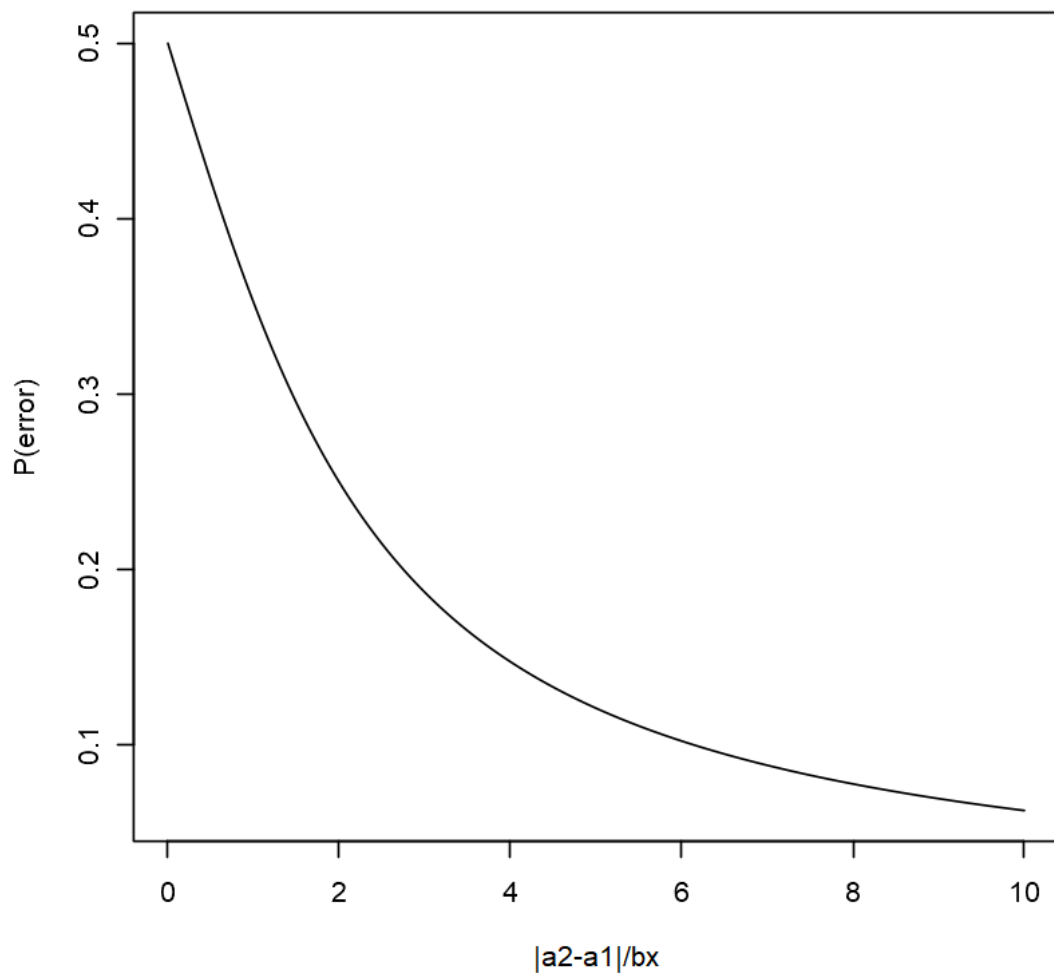
f) $P_e = \lambda_{21} P(w_2) \displaystyle\int_{-\infty}^{x_0} P(x|w_2)\, dx + \lambda_{12} P(w_1) \int_{x_0}^{+\infty} P(x|w_1)\, dx =$

$2 P(w_2) \displaystyle\int_{-\infty}^{x_0} P(x|w_2)\, dx + P(w_1) \int_{x_0}^{+\infty} P(x|w_1)\, dx$

```r
x <- seq(0, 10, length.out = 1000)
y <- 0.5 - (1/pi) * atan(x/2)

plot(x, y, type = "l", main = "P(error) plot",
     xlab = "|a2-a1|/bx", ylab = "P(error)")
```

P(error) plot

0.5 Question 5

5. We have a SVM to be trained on a set of inputs X_n, where $n = 1, \ldots, N$, together with a corresponding set of target values (t_n = +1 or - 1). The goal is to maximize the margin and at the same time allows some small misclassifications. An example of classify as follows:

$$y(\mathbf{x}_n) = \mathbf{w}^T \phi(\mathbf{x}_n) + b; \qquad \hat{t}_n = \begin{cases} +1, & \text{if } y(\mathbf{x}_n) \geq 0 \\ -1, & \text{otherwise} \end{cases}$$

We therefore minimize:

$$C \sum_{n=1}^{N} \xi_n + \frac{1}{2}\|\mathbf{w}\|^2$$

$$\text{s.t.} \qquad t_n y(\mathbf{x}_n) \geq 1 - \xi_n, \qquad\qquad n = 1, \ldots, N$$
$$\xi_n \geq 0, \qquad\qquad n = 1, \ldots, N$$

Where $C > 0$ is a parameter that controls the trade-off between the stack variable penalty and the margin. Figure below illustrates the decision boundaries for four different SVMs

using different values of C and different kernels. For each, specify which setting was used to get the result.

(Note: You should explain why you have chosen each specific choice for the corresponding figures)
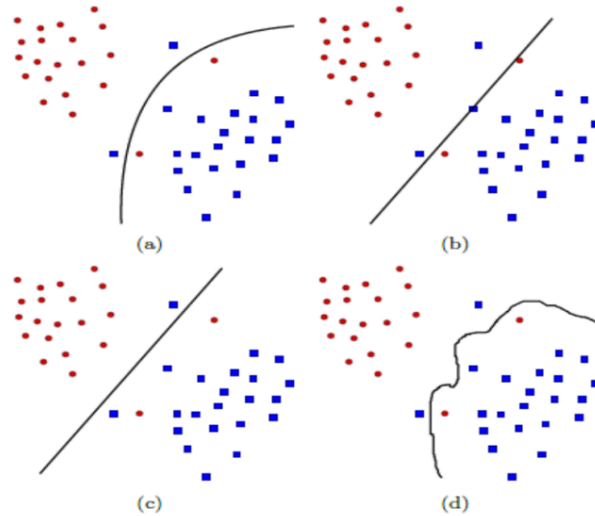
Decision boundaries corresponding to the four different SVMs.

1. $C = 1$ and no kernel is used,
2. $C = 0.1$ and no kernel is used,
3. $C = 0.1$ and the kernel is $K(\mathbf{x}_i, \mathbf{x}_j) = \exp - \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2}$.
4. $C = 0.1$ and the kernel is $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j + (\mathbf{x}_i^T \mathbf{x}_j)^2$

1.c The decision boundary is a straight line with minimal margin, focusing on classifying all training points correctly

2.b The decision boundary is still a straight line but allows some misclassifications to achieve a larger margin

3.a The decision boundary is non-linear and flexible, accommodating the data points more smoothly.

8

The margin is wider, allowing some misclassifications

4.d The decision boundary is more complex, capturing quadratic relationships between the features. It also allows some misclassifications for a larger margin