1 Chapter 8 Binary Classification

Problem 8.3 *Consider binary classification with a general class prior $p(C)$ and a Gaussian measurement distribution. Please refer to the CCSDs in section 8.2.*

(a) Derive the decision rule of the Bayesian observer in case 2 (CCSDs uniform on an interval). The answer will involve more than one cumulative standard normal distribution (Φ_{standard}), and it is not a pretty rule.

(b) Repeat for case 4 (CCSDs are equivariant Gaussians).

Answer:
(a)

$$p(C \mid x) \propto p(C) \cdot p(x \mid C)$$

To compute $p(x \mid C)$, we marginalize over s:

$$p(x \mid C = c) = \int p(x \mid s) p(s \mid C = c)\, ds = \frac{1}{a} \int_{\text{range}_c} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-s)^2}{2\sigma^2}\right) ds$$

This is a convolution of a uniform and a Gaussian distribution, and the result is:

$$p(x \mid C = c) = \frac{1}{a}\left[\Phi\left(\frac{s_{\max} - x}{\sigma}\right) - \Phi\left(\frac{s_{\min} - x}{\sigma}\right)\right]$$

where $s_{\min}$ and $s_{\max}$ are the endpoints of the uniform range:

$$\text{for } C = -1 : [-a, 0], \quad \text{for } C = +1 : [0, a]$$

Decision Rule: Report $C = +1$ if:

$$\frac{p(C = +1)}{p(C = -1)} \cdot \frac{\Phi\left(\frac{a-x}{\sigma}\right) - \Phi\left(\frac{0-x}{\sigma}\right)}{\Phi\left(\frac{0-x}{\sigma}\right) - \Phi\left(\frac{-a-x}{\sigma}\right)} > 1$$

Otherwise, report $C = -1$.

(b)
We are given that:

- $s \sim \mathcal{N}(-\mu, \sigma_s^2)$ if $C = -1$
- $s \sim \mathcal{N}(\mu, \sigma_s^2)$ if $C = +1$
- Observation: $x \sim \mathcal{N}(s, \sigma^2)$

Because both s and x are Gaussian (and the observation noise is additive), the marginal distribution over x given class C is:

$$p(x \mid C = \pm 1) = \mathcal{N}(\pm\mu, \sigma^2 + \sigma_s^2)$$

The LLR:

$$d(x) = \log \frac{p(C = +1 \mid x)}{p(C = -1 \mid x)} = \log \frac{p(C = +1)}{p(C = -1)} + \log \frac{p(x \mid C = +1)}{p(x \mid C = -1)}$$

Using the Gaussian likelihood:

$$\log \frac{p(x \mid C = +1)}{p(x \mid C = -1)} = \frac{2\mu x}{\sigma_s^2 + \sigma^2}$$

Thus, here is the decision rule: Report $C = +1$ if

$$d(x) = \log \frac{p(C = +1)}{p(C = -1)} + \frac{2\mu x}{\sigma_s^2 + \sigma^2} > 0$$

Rewriting:

$$x > -\frac{\sigma_s^2 + \sigma^2}{2\mu} \log \left(\frac{p(C = +1)}{p(C = -1)} \right)$$

From the calculation we can see that the decision rule is linear in x. The decision boundary shifts based on the prior odds and the signal-to-noise ratio. The greater the separation between the class means (larger μ), the smaller the required shift in x to make a confident decision.