CHAPTER 3:

BAYESİAN DECİSİON THEORY

Probability and Inference

- Result of tossing a coin is $\in$ {Heads,Tails}
- Random variable $X \in \{1,0\}$, where 1 = Heads, 0 = tails

  Bernoulli: $P\{X = 1\} = p_o$

  $P\{X = 0\} = (1 - p_o)$
- Sample: $\mathbf{X} = \{x^t\}_{t=1}^{N}$

  Estimation: $p_o = \#\,\{\text{Heads}\}/\#\{\text{Tosses}\} = \sum_t x^t / N$
- Prediction of next toss:

  Heads if $p_o > ½$, Tails otherwise

Classification

- Example: Credit scoring
  - Inputs are income and savings
  - Output is low-risk vs high-risk
- Input: $\boldsymbol{x} = [x_1, x_2]^T$ Output: C belongs to {0,1}
- Prediction:

$$\text{choose} \begin{cases} C = 1 & \text{if } P(C = 1 \mid x_1, x_2) > 0.5 \\ C = 0 & \text{otherwise} \end{cases}$$

or

$$\text{choose} \begin{cases} C = 1 & \text{if } P(C = 1 \mid x_1, x_2) > P(C = 0 \mid x_1, x_2) \\ C = 0 & \text{otherwise} \end{cases}$$

Bayes' Rule

$$\underbrace{P(C \mid \mathbf{x})}_{\textit{posterior}} = \frac{\overbrace{P(C)}^{\textit{prior}}\ \overbrace{p(\mathbf{x} \mid C)}^{\textit{likelihood}}}{\underbrace{p(\mathbf{x})}_{\textit{evidence}}}$$

For the case of 2 classes, C = 0 and C = 1:

$$P(C=0) + P(C=1) = 1$$

$$p(\mathbf{x}) = p(\mathbf{x} \mid C=1)P(C=1) + p(\mathbf{x} \mid C=0)P(C=0)$$

$$p(C=0 \mid \mathbf{x}) + P(C=1 \mid \mathbf{x}) = 1$$

Bayes' Rule: K>2 Classes

$$P(C_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid C_i)P(C_i)}{p(\mathbf{x})}$$

$$= \frac{p(\mathbf{x} \mid C_i)P(C_i)}{\sum_{k=1}^{K} p(\mathbf{x} \mid C_k)P(C_k)}$$

$$P(C_i) \geq 0 \text{ and } \sum_{i=1}^{K} P(C_i) = 1$$

To classify **x**: choose C_i if $P(C_i \mid \mathbf{x}) = \max_k P(C_k \mid \mathbf{x})$

Losses and Risks

- Actions: α_i : choose class C_i
- Loss of α_i when the actual class is C_k : λ_{ik}
- Expected risk (Duda and Hart, 1973)

$$R(\alpha_i \mid \mathbf{x}) = \sum_{k=1}^{K} \lambda_{ik} P(C_k \mid \mathbf{x})$$

$$\text{given } \mathbf{x}, \text{choose action } \alpha_i \text{ if } R(\alpha_i \mid \mathbf{x}) = \min_k R(\alpha_k \mid \mathbf{x})$$

Losses and Risks: 0/1 Loss

$$\lambda_{ik} = \begin{cases} 0 \text{ if } i = k \\ 1 \text{ if } i \neq k \end{cases}$$

$$\begin{aligned} R(\alpha_i \mid \mathbf{x}) &= \sum_{k=1}^{K} \lambda_{ik} P(C_k \mid \mathbf{x}) \\ &= \sum_{k \neq i} P(C_k \mid \mathbf{x}) \\ &= 1 - P(C_i \mid \mathbf{x}) \end{aligned}$$

For minimum risk, choose the most probable class

Losses and Risks: Misclassification Cost

What class C_i to pick or to Reject all classes?

Assume:

- there are K classes
- there is a loss function: cost of making a misclassification

  λ_{ik} : cost of misclassifying an instance as class C_i when it is actually of class C_k
- there is a "Reject" option (i.e., not to classify an instance in any class)

  Let the cost of "Reject" be λ.

A possible loss function is:

$$\lambda_{ik} = \begin{cases} 0 & \text{if } i = k \\ \lambda & \text{if } i = K+1, \quad 0 < \lambda < 1 \\ 1 & \text{otherwise} \end{cases}$$

For minimum risk, choose most probable class, unless is better to reject

$$\begin{array}{ll} \text{choose } C_i & \text{if } P(C_i \mid \mathbf{x}) > P(C_k \mid \mathbf{x}) \;\; \forall k \neq i \text{ and } P(C_i \mid \mathbf{x}) > 1 - \lambda \\ \text{reject} & \text{otherwise} \end{array}$$

Example: Exercise 4 from Chapter 4

Assume 2 classes: C_1 and C_2

- Case 1: Assume the two misclassifications are equally costly, and there is no reject option:

  $\lambda_{11} = \lambda_{22} = 0, \quad \lambda_{12} = \lambda_{21} = 1$

- Case 2: Assume the two misclassifications are not equally costly, and there is no reject option:

  $\lambda_{11} = \lambda_{22} = 0, \quad \lambda_{12} = 10, \quad \lambda_{21} = 5$

- Case 3: Like Case 2 but with a reject option:

  $\lambda_{11} = \lambda_{22} = 0, \quad \lambda_{12} = 10, \quad \lambda_{21} = 5, \qquad \lambda = 1$

See optimal decision boundaries on the next slide

Different Losses and Reject

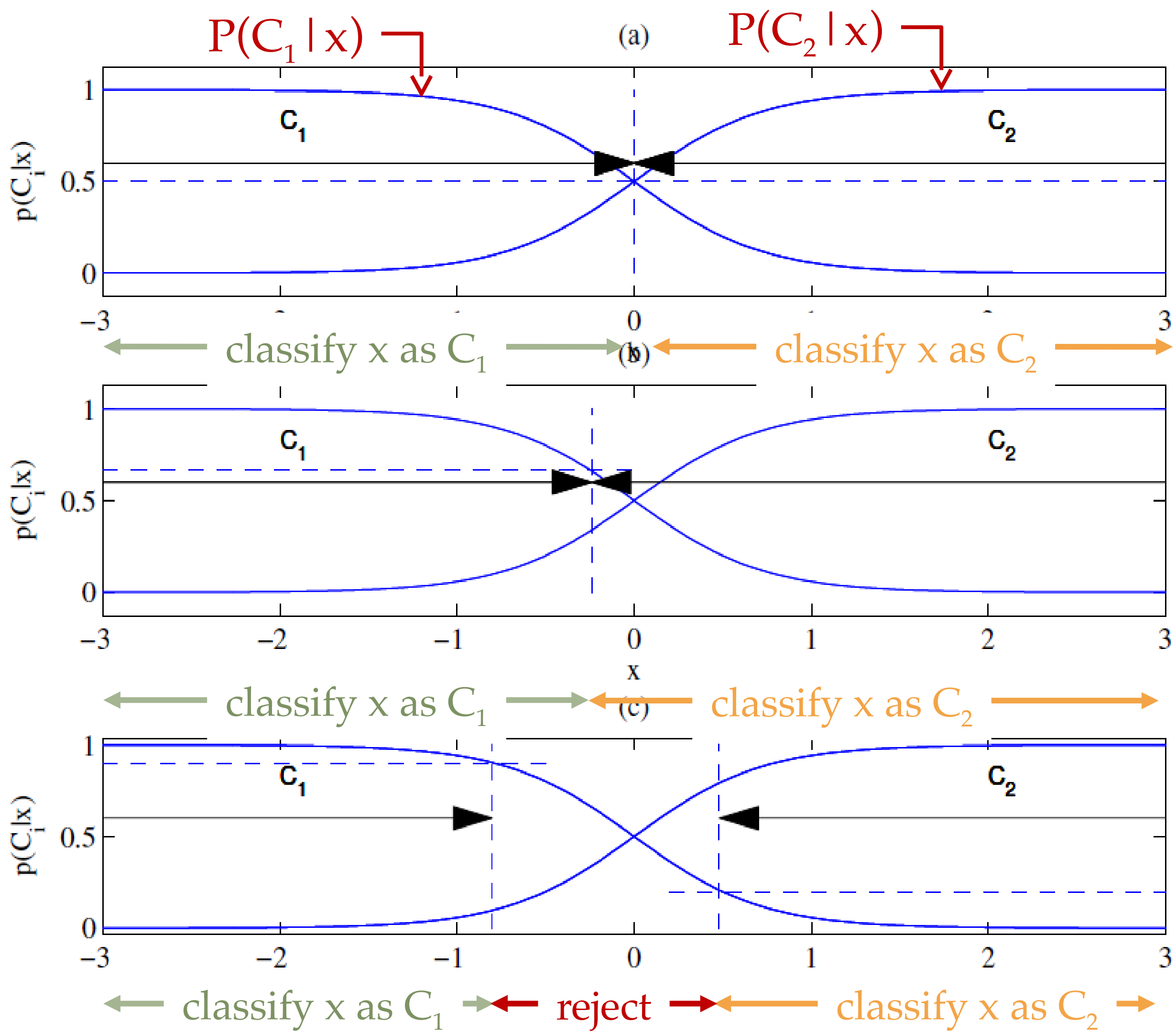
See calculations for these plots on solutions to Exercise 4

Equal losses

Unequal losses

the boundary shifts toward the class that incurs the most cost when misclassified

With reject



Discriminant Functions

Classification can be seen as implementing a set of discriminant functions $g_i(x)$:

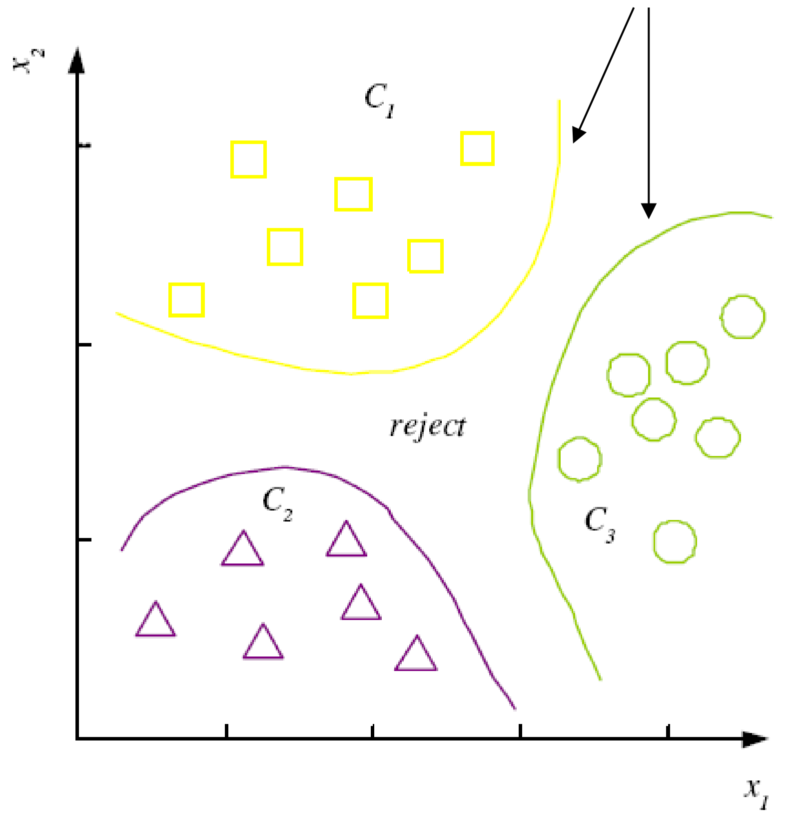
choose C_i if $g_i(\mathbf{x}) = \max_k g_k(\mathbf{x})$

$$g_i(\mathbf{x}) = P(C_i \mid \mathbf{x}) \approx p(\mathbf{x} \mid C_i) P(C_i)$$

K decision regions $R_1, \ldots, R_K$

$$R_i = \{\mathbf{x} \mid g_i(\mathbf{x}) = \max_k g_k(\mathbf{x})\}$$

$g_i(\mathbf{x}), i = 1, \ldots, K$



K=2 Classes
see Chapter 3 Exercises 2 and 3

Some alternative ways of combining discriminant functions $g_1(\mathbf{x})= P(C_1|\mathbf{x})$ and $g_2(\mathbf{x})= P(C_2|\mathbf{x})$ into just one $g(x)$:

- define $g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x})$

$$\text{choose } \begin{cases} C_1 \text{ if } g(\mathbf{x}) > 0 \\ C_2 \text{ otherwise} \end{cases}$$

- *In terms of log odds*: $\log[P(C_1|\mathbf{x})/P(C_2|\mathbf{x})]$

define $$g(\mathbf{x}) = \log \frac{P(C_1 \mid \mathbf{x})}{P(C_2 \mid \mathbf{x})}$$

$$\text{choose } \begin{cases} C_1 \text{ if } g(\mathbf{x}) > 0 \\ C_2 \text{ otherwise} \end{cases}$$

- *In terms of likelihood ratio*: $P(\mathbf{x}|C_1)/P(\mathbf{x}|C_2)$

define $$g(\mathbf{x}) = \frac{P(C_1 \mid \mathbf{x})}{P(C_2 \mid \mathbf{x})} = \frac{P(\mathbf{x} \mid C_1)}{P(\mathbf{x} \mid C_2)} \frac{P(C_1)}{P(C_2)}$$

$$\text{choose } \begin{cases} C_1 \text{ if } g(\mathbf{x}) > 1 \\ C_2 \text{ otherwise} \end{cases}$$

if the priors are equal, $P(C_1) = P(C_2)$, then the discriminant = likelihood ratio