Machine Learning 1

Lecture 5.4 - Supervised Learning
Classification - DecisionTheory

Erik Bekkers

(Bishop 1.5)

Decision theory

‣ Dataset: Input vectors $\underline{x} \in \mathbb{R}^D$, ground truth targets $t \in \{C_1, \ldots, C_k\}$

‣ Divide input space $\mathbb{R}^D$ into K decision regions $\mathcal{R}_k$, $k = 1, \ldots, K$

‣ Every observed datapoint $\begin{cases} \text{ground truth} & t_n = C_j \\ \text{prediction} & \hat{t}_n = C_k \quad (\underline{x} \in \mathcal{R}_k) \end{cases}$

‣ <mark>Confusion matrix:</mark> ground truth classes vs. predicted classes

$$
\text{ground truth} \rightarrow
\begin{array}{c}
 & \mathcal{R}_1 & \mathcal{R}_2 & \ldots & \mathcal{R}_K \\
C_1 & 6 & 1 & \ldots & 0 \\
C_2 & 5 & 3 & \ldots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
C_K & 2 & 0 & \ldots & 8
\end{array}
$$

$\leftarrow$ given by classifier $\hat{t}_n = C_k$

‣ <mark>Diagonal elements: correctly classified</mark>

‣ <mark>Off-diagonal elements: misclassified</mark>

Decision theory: Misclassification Rate

‣ Classification goal: Minimize the misclassification rate

‣ Assume observations are drawn from joint distribution $p(\underline{x}, t)$

‣ Probability of a misclassification:

$$p(\text{mistake}) = \sum_{i=1}^{K} \sum_{k \neq i} p(\mathbf{x} \in R_i, C_k)$$

$$= 1 - \sum_{k=1}^{K} p(x \in R_k, C_k)$$

Minimizing misclassification rate

‣ Assign x to class C_k if $\quad p(\underline{x}, t = C_k) > p(\underline{x}, t = C_j), \ j \neq k$

‣ Note: $p(x, C_k) = \underline{p(C_k|x)}p(x)$

Check for the largest post. class prob

$$p(C_k|x) > p(C_j|x), \ j \neq k$$

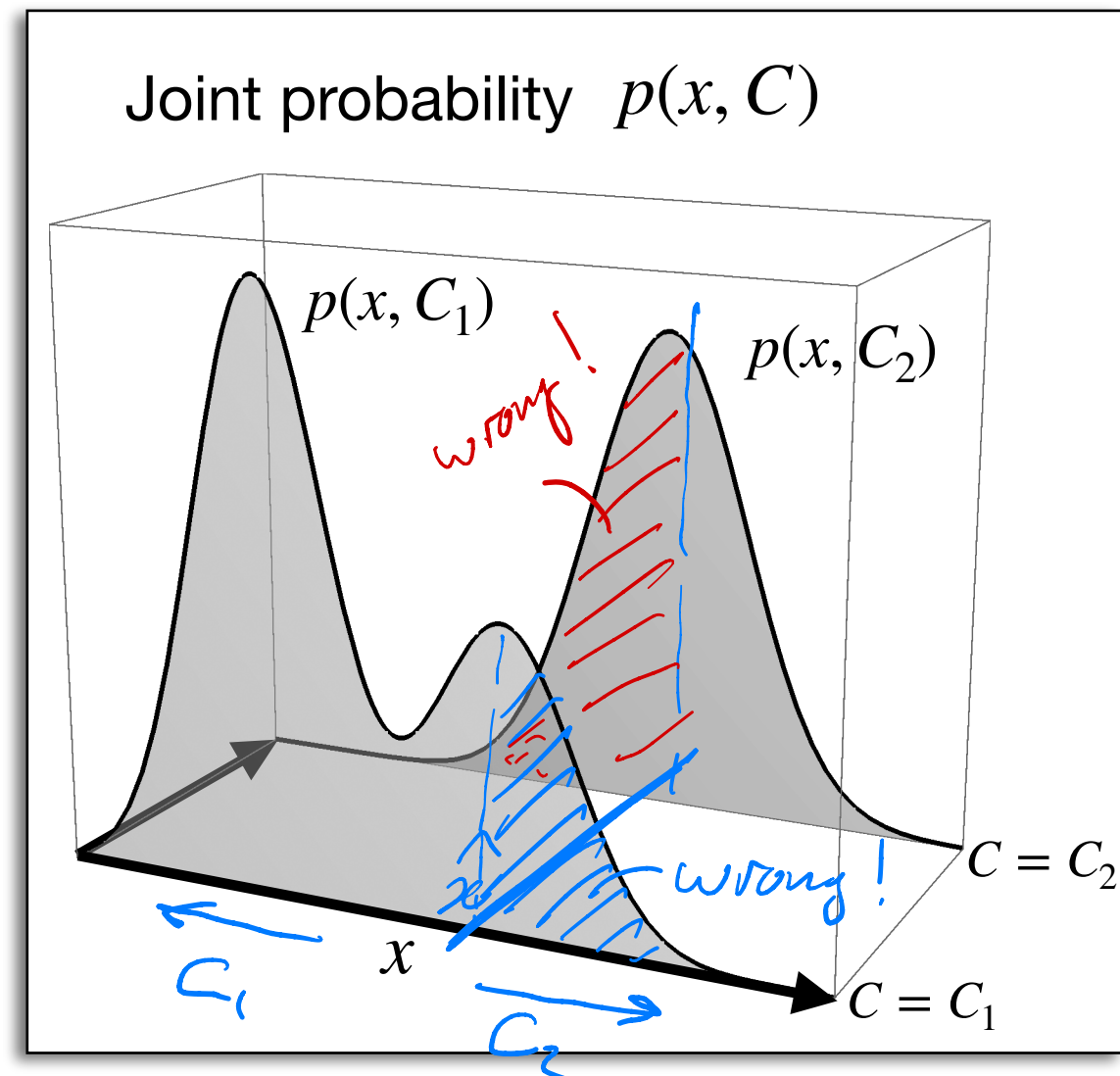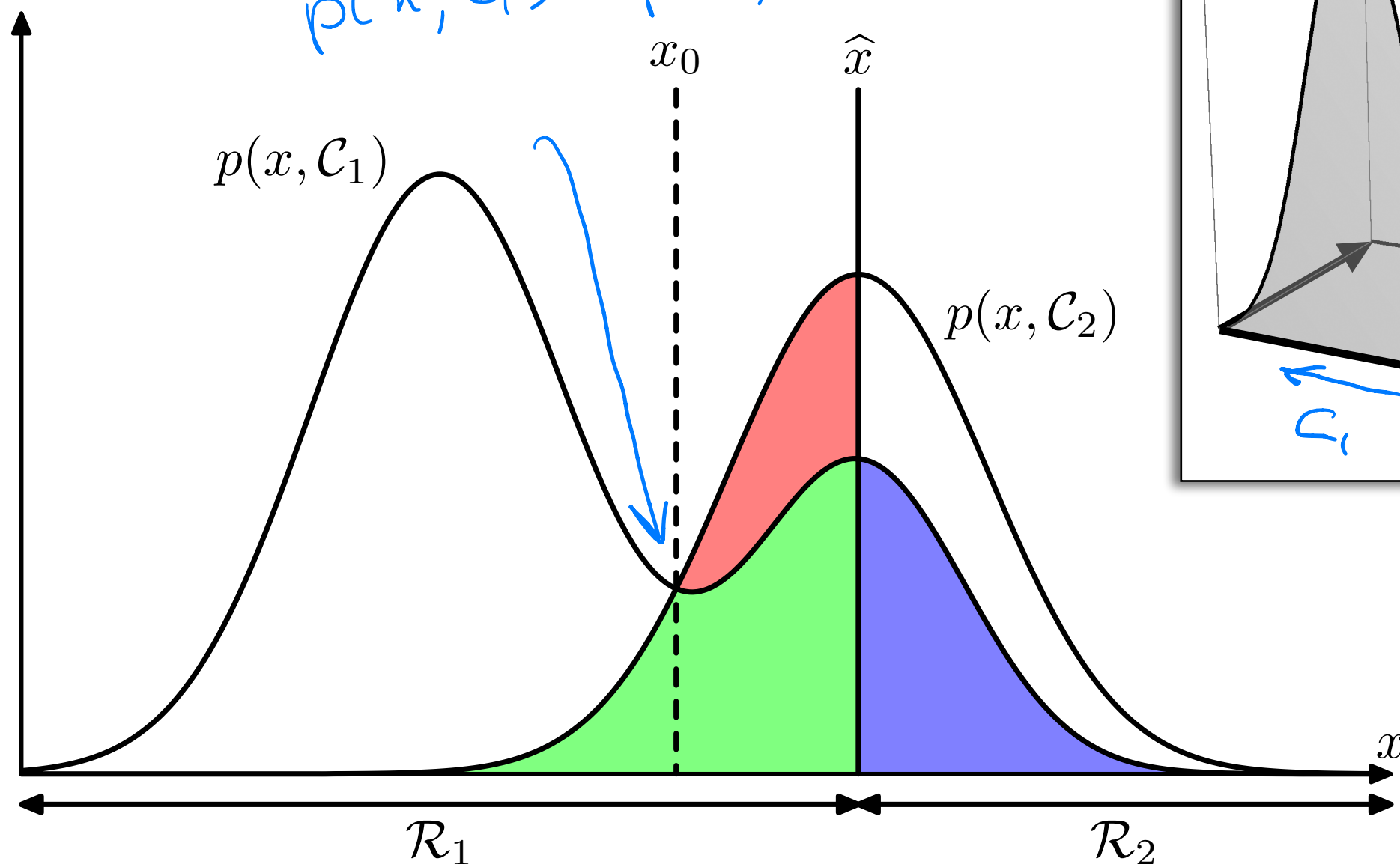# Decision theory: Misclassification Rate

Figure: joint probability distributions and decision boundary (Bishop 1.24)

Minimizing the Misclassification Rate: Problems

① ‣ Not all errors have the same impact!

Example: Medical diagnosis of cancer

‣ Error 1: Label a healthy person as having cancer.

‣ Error 2: Label a sick person as healthy. Lack of treatment!

‣ If cancer only occurs in 1% of all patients, a classifier which labels everyone as healthy has a misclassification rate of 1%!

② Class Imbalance

Expected Loss

‣ Possible solution: use different weights for different error types

$$L = \begin{pmatrix} \overset{\text{label cancer}}{0} & \overset{\text{label healthy}}{1000} \\ 1 & 0 \end{pmatrix} \begin{array}{l} \text{true cancer} \\ \text{true healthy} \end{array}$$

‣ Expected loss: $\mathbb{E}[L] = \sum_{k,j} L_{kj} \int_{\mathcal{R}_j} p(x, C_k)\mathrm{d}x$

errors $0 \neq k$

Minimize expected loss:

‣ Assign x to C_k if $\sum_{j=1}^{K} L_{jk} p(x, C_j)$ is minimal

Classification Strategies

① ‣ *Discriminant functions*

Direct mapping of input to target $\quad t = y(\underline{x}, \underline{w})$

② ‣ *Probabilistic discriminative models*

Posterior class probabilities: $\quad p(C_k | x)$

③ ‣ *Probabilistic generative models*

Class-conditional densities: $\quad p(x | C_k)$

Prior class probabilities: $\quad p(C_k)$

1. $p(x, C_k) = p(x | C_k)\, p(C_k)$

2. $p(C_k | x) = \dfrac{p(x | C_k)\, p(C_k)}{p(x)}$