Bayesian Decision Theory

Fall 2019

Instructor: Shandian Zhe
zhe@cs.utah.edu
School of Compputing

THE UNIVERSITY OF UTAH

Given $\boldsymbol{x}$, we want to predict $\boldsymbol{t}$

Cancer
Stock price
Weather
….

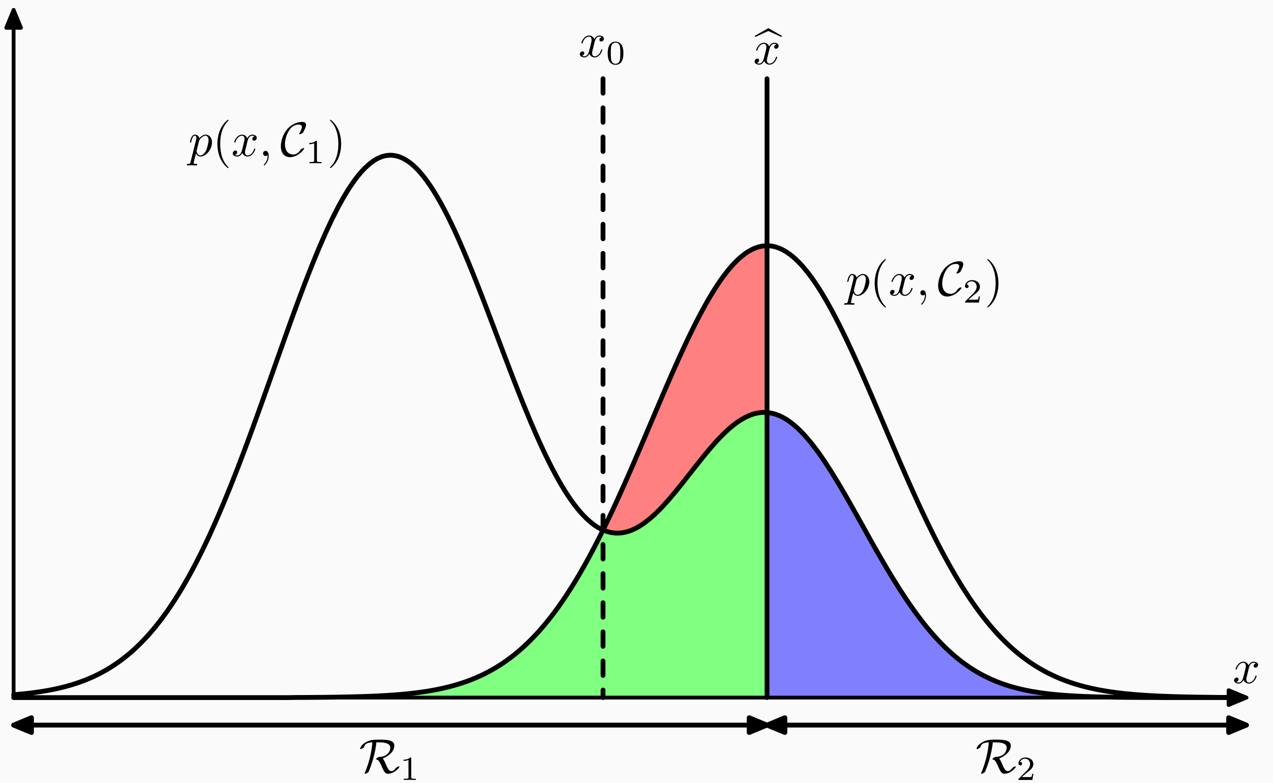
- Inference step
  - Determine either $p(\boldsymbol{t}|\boldsymbol{x})$ or $p(\boldsymbol{x},\boldsymbol{t})$ (from training data)

- Decision Step
  - For Given $\boldsymbol{x}$, determine optimal $\boldsymbol{t}$

Let us first consider the classification problem

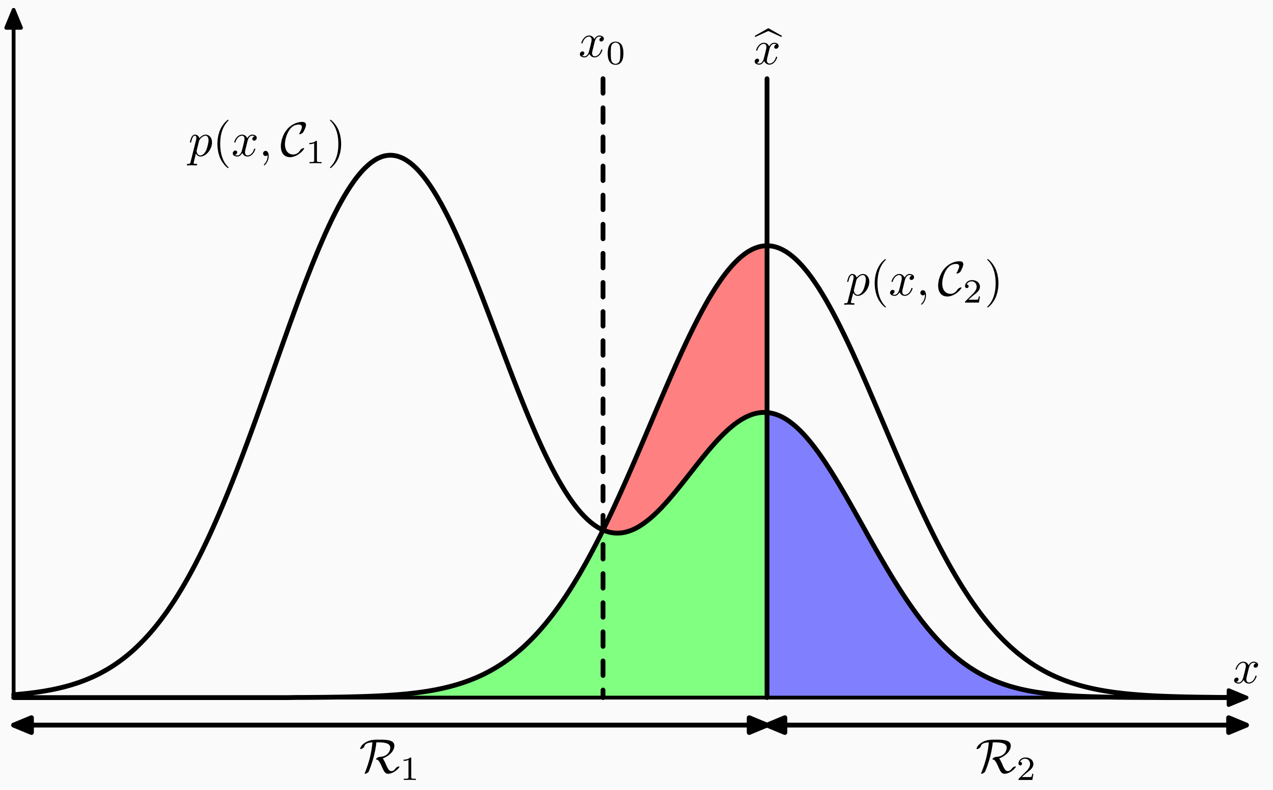
- $\boldsymbol{t} \in \{C_1, \ldots, C_K\}$
- Decision regions R_k: if $\boldsymbol{x}$ falls in , predict C_k
- Decision boundaries/surfaces: boundaries between different decision regions

Minimum misclassification rate



$$
\begin{aligned}
p(\text{mistake}) &= p(\mathbf{x} \in \mathcal{R}_1, \mathcal{C}_2) + p(\mathbf{x} \in \mathcal{R}_2, \mathcal{C}_1) \\
&= \int_{\mathcal{R}_1} p(\mathbf{x}, \mathcal{C}_2)\,\mathrm{d}\mathbf{x} + \int_{\mathcal{R}_2} p(\mathbf{x}, \mathcal{C}_1)\,\mathrm{d}\mathbf{x}.
\end{aligned}
$$

Minimum misclassification rate



Question: where shall we set the decision boundary to minimize the misclassification rate? Why?

Minimum misclassification rate

- In general for *K* classes

$$
\begin{aligned}
p(\text{correct}) &= \sum_{k=1}^{K} p(\mathbf{x} \in \mathcal{R}_k, \mathcal{C}_k) \\
&= \sum_{k=1}^{K} \int_{\mathcal{R}_k} p(\mathbf{x}, \mathcal{C}_k)\, \mathrm{d}\mathbf{x}
\end{aligned}
$$

$$\boxed{p(\mathcal{C}_k|\mathbf{x})}\, p(\mathbf{x})$$

How to find regions that maximize the probability of correctness?

Minimum misclassification rate

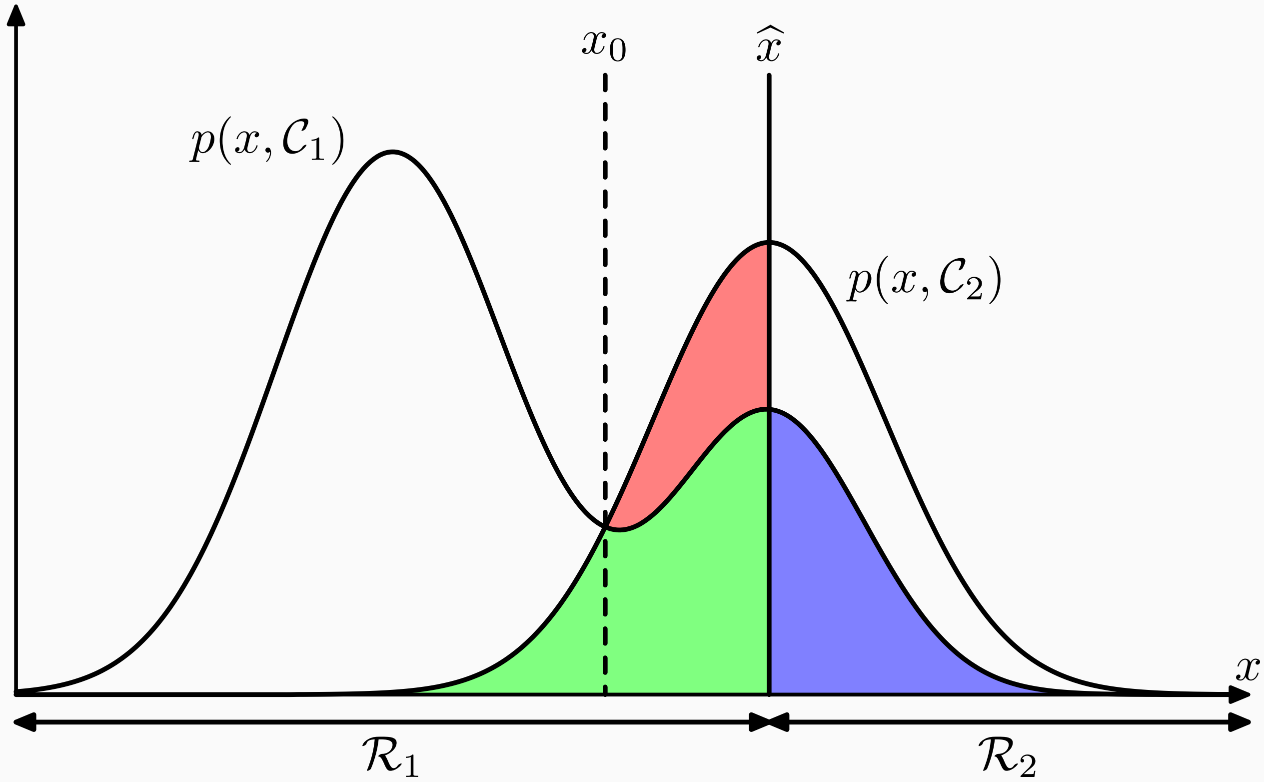
- In general for *K* classes

$$
\begin{aligned}
p(\text{correct}) &= \sum_{k=1}^{K} p(\mathbf{x} \in \mathcal{R}_k, \mathcal{C}_k) \\
&= \sum_{k=1}^{K} \int_{\mathcal{R}_k} p(\mathbf{x}, \mathcal{C}_k)\, \mathrm{d}\mathbf{x}
\end{aligned}
$$

$$
p(\mathcal{C}_k|\mathbf{x})p(\mathbf{x})
$$

Each ***x*** should be assigned the class having the largest posterior probability $p(C_k|\boldsymbol{x})$

Look back


x_0
$\widehat{x}$
$p(x, \mathcal{C}_1)$
$p(x, \mathcal{C}_2)$
x
$\mathcal{R}_1$
$\mathcal{R}_2$

Minimum Expected Loss

- In practice, mistakes in predicting different classes may lead to different costs

Example: classify medical images as 'cancer' or 'normal'

Truth \ Decision	cancer	normal
cancer	0	1000
normal	1	0

Minimum Expected Loss

- Define a *cost function*, associate the cost of classifying *k* to *j* with L_{kj}

$$\mathbb{E}[L] = \sum_k \sum_j \int_{\mathcal{R}_j} L_{kj} p(\mathbf{x}, \mathcal{C}_k) \, \mathrm{d}\mathbf{x}.$$

- We want to find the decision regions R_j that minimize the expected loss

Minimum Expected Loss

- Define a *cost function*, associate the cost of classifying *k* to *j* with L_{kj}

$$\mathbb{E}[L] = \sum_k \sum_j \int_{\mathcal{R}_j} L_{kj} p(\mathbf{x}, \mathcal{C}_k) \, \mathrm{d}\mathbf{x}.$$

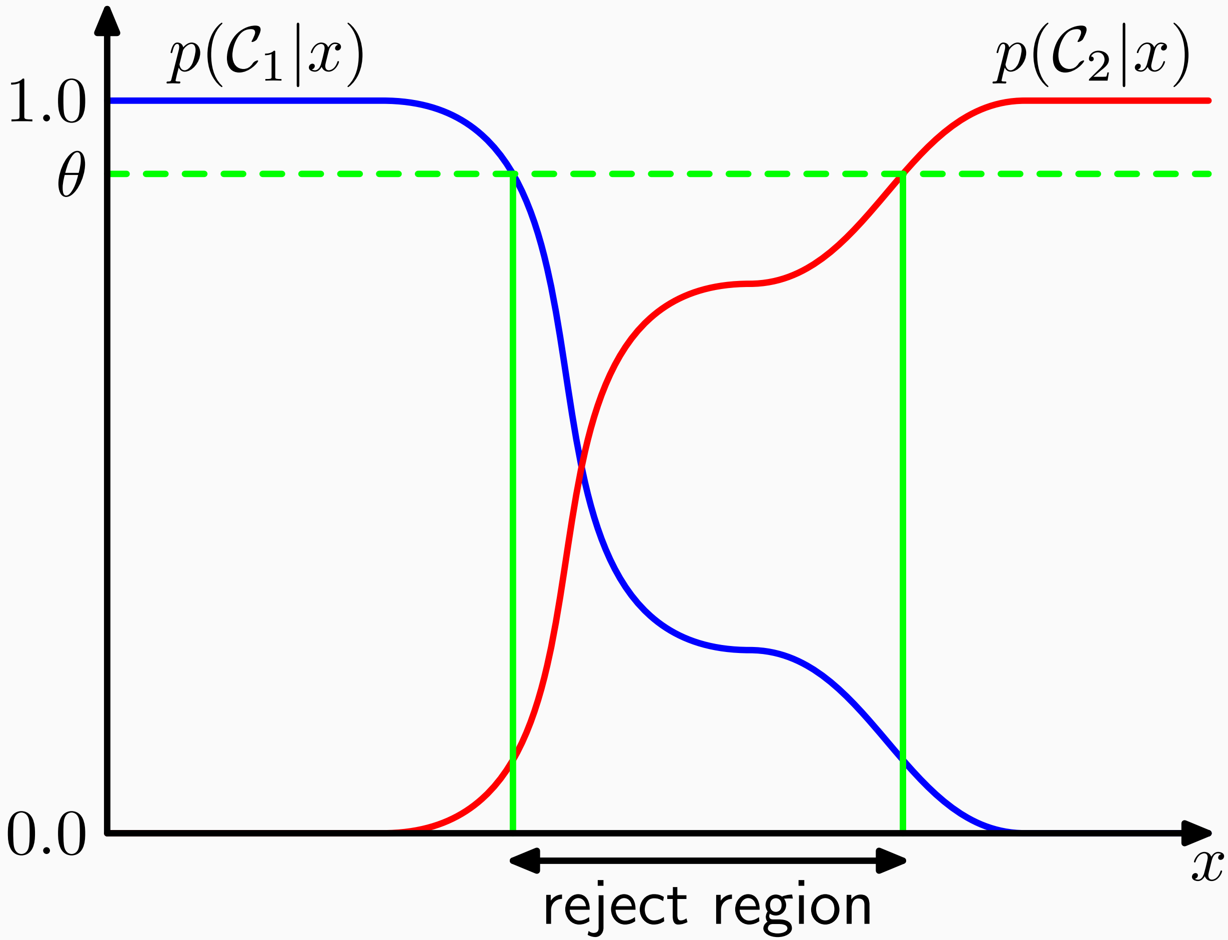
- Rule: Assign each ***x*** to the class for which

$$\sum_k L_{kj} p(\mathcal{C}_k | \mathbf{x})$$

is a minimum

Reject option

- When the largest posterior probability is still too small



Decision for continuous variables

- Inference step
  - Determine $p(\mathbf{x},t)$

- Decision step
  - For any given $\mathbf{x}$, make optimal prediction $y(\mathbf{x})$ for t

Decision for continuous variables

- Inference step
  - Determine $p(\mathbf{x},t)$

- Decision step
  - For any given $\mathbf{x}$, make optimal prediction $y(\mathbf{x})$ for t

Loss function

$$\mathbb{E}[L] = \iint L(t, y(\mathbf{x}))p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$$

The squared loss function

Minimize $$\mathbb{E}[L] = \iint \{y(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$$

The squared loss function

Minimize $$\mathbb{E}[L] = \iint \{y(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\, \mathrm{d}\mathbf{x}\, \mathrm{d}t$$

$$\Downarrow$$

$$y(\mathbf{x}) = \mathbb{E}[t|\mathbf{x}]$$

What you need to know

- What is the decision? What is the difference between the decision and inference?
- How to find optimal decision regions for classification?
- How to find optimal decisions for continuous variables?