5. Bayesian Decision Theory

See: Duda and Hart Chapter 2.

5.1 The Bayes Classifier

Classifying Fish

- Simple model:
  - No posterior knowledge (i.e. no measurements)
  - Two classes

    ω_1 =“sea bass”

    ω_2=“salmon”
  - Given: $P(\omega_1)$ and $P(\omega_2)$
  - Goal:
    - Minimize the number of fish that get the wrong label

How would you set up a decision rule?

Classifying Fish

Sea bass

Salmon

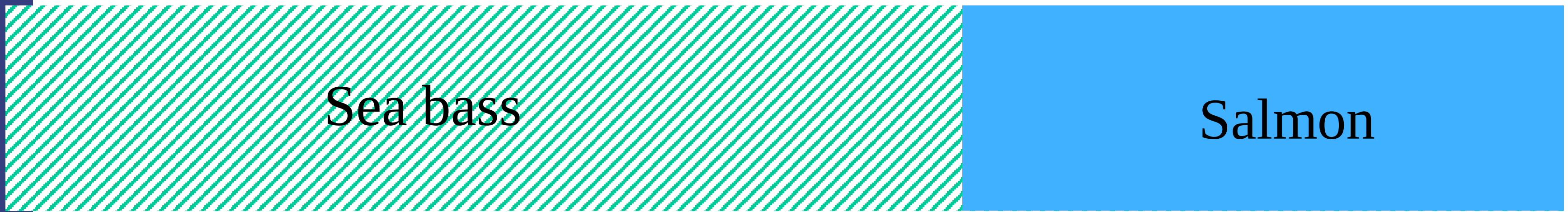
$P(\omega_1)$

$P(\omega_2)$

Classify every fish as

Classifying Fish

Incorrectly classified

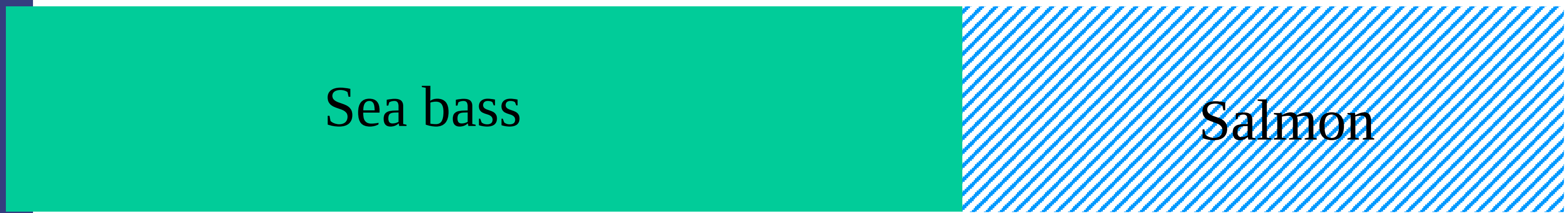


$P(\omega_1)$ $P(\omega_2)$

Classify every fish as salmon

Classifying Fish

Incorrectly classified



$P(\omega_1)$ $P(\omega_2)$

Classify every fish as “see bass”

Smaller number of fish with wrong label

Generalization

- Minimize number of wrong labels

  $\mapsto$pick class with highest probability

Formal notation:

$$\overline{\omega_i} = \arg\max_{\omega_k} P(\omega_k)$$

Available Measurements x

- Feature vector x from measurement
- Probabilities depend on x

$$P(\omega_k \mid x)$$

- Definition conditional probability:

$$P(\omega_k \mid x) = \frac{P(\omega_k, x)}{P(x)}$$

Bayes Decision Rule: Draft Version

- Bayes decision rule

$$\overline{\omega_i} = \arg\max_{\omega_k} P(\omega_k \mid x)$$

Ugly: usually x is measured for a given class ω_k

Rewrite Bayes Decision Rule

$$\overline{\omega_i} = \arg\max_{\omega_k} P(\omega_k \mid x)$$

$$= \arg\max_{\omega_k} \frac{P(x \mid \omega_k)P(\omega_k)}{P(x)}$$

$$= \arg\max_{\omega_k} P(x \mid \omega_k)P(\omega_k)$$

Use definition of cond. probability

$$P(\omega_k \mid x) = \frac{P(\omega_k, x)}{P(x)}$$

$$= \frac{P(x \mid \omega_k)P(\omega_k)}{P(x)}$$

P(x) does not affect decision

Bayes Decision Rule

$$\overline{\omega_i} = \arg\max_{\omega_k} P(x \mid \omega_k) P(\omega_k)$$

Terminology

Prior: $P(\omega_k)$

Posterior: $P(\omega_k \mid x)$

Cost of Making Errors

- The fish is a “salmon”
- You classify it as a “sea bass”
- You sell it as a “sea bass”

$\mapsto$ angry customer

Cost Making Errors

- The fish is a “sea bass”
- You classify it as a “salmon”
- You sell it as a “salmon”

$\mapsto$ lost revenue

Loss Function

		Fish is a	
		Sea bass	Salmon
Sold as	Sea bass	0$	2$
	Salmon	1$	0$

Loss Function and Conditional Risk

- True classes $\{\omega_1, \omega_2, \ldots, \omega_c\}$
- Actions taken $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$
- Loss function $\lambda(\alpha_i \mid \omega_j)$
- Conditional risk

$$R(\alpha_i \mid x) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

How to include p(x)
to estimate overall loss/risk?

Overall Risk

- Decision rule: map feature vector to action
  - $x \mapsto \alpha$
- Goal:

  Determine decision rule that minimizes overall risk:

$$R = \int R(\alpha(x) \mid x)\, p(x)\, dx$$

$\mapsto$ to minimize R, pick the action that minimizes the conditional risk for a specific x

Example: two-class problem (1)

- Classes: ω_1, ω_2
- Actions: $\alpha_{1,}$ α_2
- For simplicity: loss: $\lambda_{ij}=\lambda(\alpha_i| \omega_j)$
- Conditional risk:

$$R(\alpha_1 \mid x) = \lambda_{11} P(\omega_1 \mid x) + \lambda_{12} P(\omega_2 \mid x)$$
$$R(\alpha_2 \mid x) = \lambda_{21} P(\omega_1 \mid x) + \lambda_{22} P(\omega_2 \mid x)$$

Example: two-class problem (2)

- Example actions
  - α_1: decide that the class is ω_1
  - α_2: decide that the class is ω_2
- decide that the class is ω_1 if:

$$R(\alpha_1 \mid x) < R(\alpha_2 \mid x) \quad \Rightarrow$$

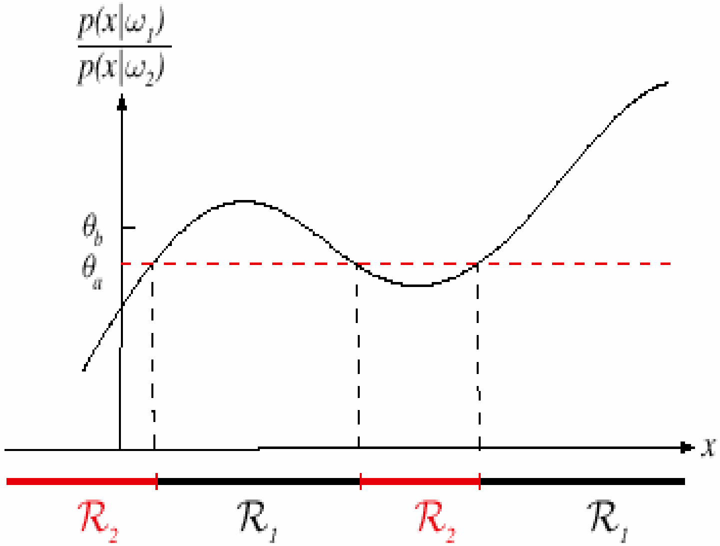
$$\lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x) < \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x) \quad \Rightarrow$$

$$(\lambda_{12} - \lambda_{22})P(\omega_2 \mid x) < (\lambda_{21} - \lambda_{11})P(\omega_1 \mid x)$$

Replaces Bayes decision rule

Example: two-class problem (3)

- Rephrase: $$\frac{P(x\,|\,\omega_1)}{P(x\,|\,\omega_2)} > \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$



$\mapsto$ tune threshold θ to tune overall risk (loss)

Minimum Error Rate Classification

General case difficult to handle
Important special case: minimze the number of errors

Actions:

α_i: decide that the class is ω_i

"Zero-one-loss"-function

$$\lambda(a_i \mid \omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \quad i, j = 1, \ldots, c$$

Conditional Risk for zero-one Loss Function

$$R(\alpha_i \mid x) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

$$= \sum_{j=1, i \neq j}^{c} P(\omega_j \mid x)$$

How can you simplify this?

Def. of zero-one loss function

$$= 1 - P(\omega_i \mid x)$$

Normalization of probability

Minimum Error Rate/ Bayes Decision Rule

- Pick i that minimizes risk:

$$R(\alpha_i \mid x) = 1 - P(\omega_i \mid x)$$

$\mapsto$ pick i that maximizes conditional probability

$$P(\omega_i \mid x)$$

$\mapsto$ Bayes decision rule

Example: two-class problem (3)

- Minimum error rate applied to example
- Action α_1: decide that the class is ω_1
- Take this action if

$$(\lambda_{12} - \lambda_{22}) P(\omega_2 \mid x) < (\lambda_{21} - \lambda_{11}) P(\omega_1 \mid x) \quad \Rightarrow$$

$$P(\omega_2 \mid x) < P(\omega_1 \mid x)$$

$\mapsto$ Recover Bayes Decision Rule

Summary 5.1. The Bayes Classifier

- Bayes classifier

$$\overline{\omega_i} = \arg\max_{\omega_k} P(x \mid \omega_k) P(\omega_k)$$

- Minimizes number of classification errors
- Generalization: minimize loss ("risk")