Announcements

- HW0 due on Thursday
- Readings up on eReserves
- The project

Pattern classification

Biometrics

CSE 190

Lecture 3

Bayesian Decision Theory
Continuous Features
(Sections 2.1-2.2)

Introduction

- The sea bass/salmon example

  - State of nature, prior
    - State of nature is a random variable
    - The catch of salmon and sea bass is equiprobable
    - $P(o_1), P(o_2)$ Prior probabilities
    - $P(o_1) = P(o_2)$ (uniform priors)
    - $P(o_1) + P(o_2) = 1$ (exclusivity and exhaustivity)
• Decision rule with only the prior information
  • Decide $\omega_1$ if $P(\omega_1) > P(\omega_2)$ otherwise decide $\omega_2$

• Use of the class–conditional information
  • $P(x \mid \omega_1)$ and $P(x \mid \omega_2)$ describe the difference in lightness between populations of sea-bass and salmon

• Posterior, likelihood, evidence
  • $P(\omega_j \mid x) = \frac{P(x \mid \omega_j) P(\omega_j)}{P(x)}$ (BAYES RULE)
  • In words, this can be said as:
    Posterior = (Likelihood * Prior) / Evidence
  • Where in case of two categories
    $$P(x) = \sum_{j=1}^{2} P(x \mid \omega_j) P(\omega_j)$$

• Since decision rule is optimal for each feature value $X$, there is no better rule for all $x$. 

• Intuitive decision rule given the posterior probabilities:
  Given $x$:
  - if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ True state of nature = $\omega_1$
  - if $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ True state of nature = $\omega_2$

Why do this?: Whenever we observe a particular $x$, the probability of error is:
  - $P(error \mid x) = P(\omega_1 \mid x)$ if we decide $\omega_2$
  - $P(error \mid x) = P(\omega_2 \mid x)$ if we decide $\omega_1$

The Maximum A Posteriori (MAP) decision rule is

$$\hat{\omega} = \arg\max_{\omega_j} P(\omega_j \mid x)$$

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

Bayes Rule

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

Because $P(x)$ is not a function of $\omega_j$

$P(x \mid \omega_j) P(\omega_j)$ is just a scaled version $P(x \mid \omega_j)$
Bayesian Decision Theory – Continuous Features

Generalization of the preceding ideas

- Use of more than one feature
- Use more than two states of nature
- Allowing actions and not only decide on the state of nature
- Introduce a loss of function (more general than the probability of error)
- Allowing actions other than classification primarily allows the possibility of rejection
- Refusing to make a decision in close or bad cases!
- Letting loss function state how costly each action taken is

Example

What is the Expected Loss for action $\alpha_i$?

For any given $x$ the expected loss for action $\alpha_i$ is

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

$R(\alpha_i | x)$ is called the Conditional Risk (or Expected Loss)

Overall risk

$$R = \int R(\alpha(x) | x) P(x) dx$$

Minimizing $R$ $\iff$ Minimizing $R(\alpha_i | x)$ for $i = 1, \ldots, a$

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

for $i = 1, \ldots, a$

Given a measured feature vector $x$, which action should we take?

Select the action $\alpha_i$ for which $R(\alpha_i | x)$ is minimum

$$\hat{\alpha}(x) = \arg \min_{\alpha_i} R(\alpha_i | x)$$

$$= \arg \min_{\alpha_i} \sum_{j=1}^{c} \lambda(\alpha_i | \omega_j) P(\omega_j | x)$$
Example: Two-Category Classification

\( \alpha_1 \): deciding \( \omega_1 \)

\( \alpha_2 \): deciding \( \omega_2 \)

\( \lambda_{ij} = \lambda_j(\alpha_i | \omega_j) \)

loss incurred for deciding \( \omega_i \) when the true state of nature is \( \omega_j \)

Conditional risk:

\[
R(\alpha_1 | x) = \lambda_{11}P(\omega_1 | x) + \lambda_{12}P(\omega_2 | x)
\]

\[
R(\alpha_2 | x) = \lambda_{21}P(\omega_1 | x) + \lambda_{22}P(\omega_2 | x)
\]

Our rule is the following:

if \( R(\alpha_1 | x) < R(\alpha_2 | x) \)

\( \lambda_{11}P(\omega_1 | x) + \lambda_{12}P(\omega_2 | x) < \lambda_{21}P(\omega_1 | x) + \lambda_{22}P(\omega_2 | x) \)

action \( \alpha_2 \): “decide \( \omega_2 \)” is taken

This results in the equivalent rule:

Decide \( \omega_1 \) if:

\[
(\lambda_{21} - \lambda_{11}) P(x | \omega_1) P(\omega_1) > (\lambda_{12} - \lambda_{22}) P(x | \omega_2) P(\omega_2)
\]

Decide \( \omega_2 \) otherwise

On to higher dimensions!