Discrete and Continuous Random Variables

- Probability mass function
  - Example
    - Bernoulli, Binomial distributions
- Probability density function
  - Example
    - Normal distribution

A Big Picture of Learning

- Supervised Learning
  - Classification, regression
  - Generalization
- Unsupervised learning
  - Clustering
  - Retrieval
- In between

Bayesian Classifiers

- General framework
  - Each data object is characterized by
    - Nature of the object \( \omega \)
    - Feature vector \( x \)
- Binary classification
  - Two possible states of nature \( \{\omega_1, \omega_2\} \)
  - Class prior probability \( P(\omega_1), P(\omega_2) \)
  - Class conditional probability \( p(x | \omega_1), p(x | \omega_2) \)
Problem Formulation

• Given prior probabilities and class conditional p.d.f
\[ p(x|\omega_1), p(\omega_1) \]
\[ p(x|\omega_2), p(\omega_2) \]
design a decision rule such that
\[ P(\text{error}) \]
is minimized

The Probability of Error

• Observe \( x \), the probability of error is
\[ P(\text{error} | x) = \begin{cases} P(\omega_1 | x) & \text{if we decide } \omega_1 \\ P(\omega_2 | x) & \text{if we decide } \omega_2 \end{cases} \]
• The average probability
\[ P(\text{error}) = \int P(\text{error} | x) p(x) dx = \int P(\text{error} | x) p(x) dx \]

Bayes Decision Rule

• A fact
\[ P(\text{error}) \geq \int \min\{P(\omega_1 | x), P(\omega_2 | x)\} p(x) dx \]
• Bayes decision rule
  Decide \( \omega_1 \) if \( P(\omega_1 | x) > P(\omega_2 | x) \);
  otherwise decide \( \omega_2 \)
Bayes Decision Rule

- Posterior class probability
  \[
  P(\omega_j | x) = \frac{p(x | \omega_j)P(\omega_j)}{p(x)}
  \]
  \[
  p(x) = p(x | \omega_1) + p(x | \omega_2)
  \]
  \[
  = p(x | \omega_1)P(\omega_1) + p(x | \omega_2)P(\omega_2)
  \]

Class-Conditional p.d.f.

Example
- Posterior (\( P(\omega_1) = 2/3, P(\omega_2) = 1/3 \))
Bayes Decision Rule

• An equivalent decision rule
  Decide $\omega_1$ if $p(x | \omega_1)P(\omega_1) > p(x | \omega_2)P(\omega_2)$
  otherwise decide $\omega_2$

A Generalized Formulation

• Allowing more than one feature
• Allowing more than two states of nature
• Allowing more than two decisions
• Introducing a loss function more general than the probability of error

Problem Formulation

• Feature vector $\bar{x}$
• $d$-dimensional feature space $\mathbb{R}^d$
• Assume complete statistical information
  – Finite set of $c$ states $\omega \in \{\omega_1, ..., \omega_c\}$
  – Class prior probability $P(\omega_j)$
  – Class conditional probability $p(\bar{x} | \omega_j)$
Problem Formulation

• Action set \( A = \{\alpha_1, \ldots, \alpha_c\} \)
• Loss function \( \ell(\alpha_i | \omega_j) \)
• Expected risk
\[
R(\alpha_i | x) = \sum_{j=1}^{c} A(\alpha_i | \omega_j) P(\omega_j | x)
\]
• Overall risk for a general decision rule \( \alpha(x) \)
\[
R(\alpha) = \int R(\alpha(x) | x) p(x) dx
\]

Bayes Risk

• A fact
\[
R(\alpha) = \int R(\alpha(x) | x) p(x) dx
\]
\[
\geq \int \min_{i \neq j} R(\alpha_i | x) p(x) dx
\]
\[
= \text{Bayes risk}
\]
• Bayes decision rule
Decide \( \alpha_i \) if \( R(\alpha_i | x) < R(\alpha_j | x), j \neq i \)

Compute Conditional Risk

\[
R(\alpha_i | x) = \sum_{j=1}^{c} A(\alpha_i | \omega_j) P(\omega_j | x)
\]
• Posterior probability
\[
P(\omega_j | x) = \frac{p(x | \omega_j) P(\omega_j)}{p(x)}
\]
\[
p(x) = \sum_{j=1}^{c} p(x | \omega_j) P(\omega_j)
\]