CS559 Machine Learning

Bayesian Decision Theory

Tian Han

Department of Computer Science
Stevens Institute of Technology

Week 2

Outline

- Introduction
- Bayesian Decision Theory
- Minimum Error Rate Classification
- Classifier and Discriminant Functions
- Three Approaches for Decision Problem

Introduction

Why and what is Bayesian decision?

- (From the Economist 2000) The essence of the Bayesian approach is to provide a mathematical rule explaining how you should change your existing beliefs in the light of new evidence.
- It allows the scientist to combine new data with their existing knowledge.
- Bayesian decision theory uses Bayes approach to analysis the problem of pattern classification.
- Quantify the trade-offs between various decisions using probability and the cost that accompany such decisions.

Assumption:

- Decision problem is posed in **probabilistic** terms.
- All of the relevant probabilities are **known**.

Fish Example

Salmon

Sea Bass

Figure: From J.Corso slides

- Classify fish as either Salmon or Sea Bass.
- Random variable ω describe the fish category. (State of nature)
  - $\omega = \omega_1$: Sea Bass
  - $\omega = \omega_2$: Salmon
- Only two fish categories.

Prior Probability

- The Prior probability reflects our prior knowledge of how likely we expect an outcome of an event **before** we actually observed such event.
- For fish example, represents how likely we are to get a sea bass or salmon before we see the next fish on the conveyor belt.
- Prior comes from prior knowledge, **NO** data have been seen yet.
- Prior might be different depending on the situation.
- If have reliable prior knowledge, USE IT!

Decision Rule based on ONLY Prior

- $P(\omega = \omega_1)$, or $P(\omega_1)$ for prior next is the sea bass.
- $P(\omega = \omega_2)$, or $P(\omega_2)$ for prior next is the salmon.

Decision Rule based on ONLY Prior

- $P(\omega = \omega_1)$, or $P(\omega_1)$ for prior next is the sea bass.
- $P(\omega = \omega_2)$, or $P(\omega_2)$ for prior next is the salmon.
- $P(\omega_1) + P(\omega_2) = 1$: either ω_1 or ω_2 must occur.

Decision Rule based on ONLY Prior

- $P(\omega = \omega_1)$, or $P(\omega_1)$ for prior next is the sea bass.
- $P(\omega = \omega_2)$, or $P(\omega_2)$ for prior next is the salmon.
- $P(\omega_1) + P(\omega_2) = 1$: either ω_1 or ω_2 must occur.
- A decision rule prescribes what actions to take based on observed data.

Decision Rule based on ONLY Prior

- $P(\omega = \omega_1)$, or $P(\omega_1)$ for prior next is the sea bass.
- $P(\omega = \omega_2)$, or $P(\omega_2)$ for prior next is the salmon.
- $P(\omega_1) + P(\omega_2) = 1$: either ω_1 or ω_2 must occur.
- A decision rule prescribes what actions to take based on observed data.
- Assume **only prior** available and **equal cost** for incorrect classifications.

Decision Rule based on ONLY Prior

- $P(\omega = \omega_1)$, or $P(\omega_1)$ for prior next is the sea bass.
- $P(\omega = \omega_2)$, or $P(\omega_2)$ for prior next is the salmon.
- $P(\omega_1) + P(\omega_2) = 1$: either ω_1 or ω_2 must occur.
- A decision rule prescribes what actions to take based on observed data.
- Assume **only prior** available and **equal cost** for incorrect classifications.
  - Decide ω_1 if $P(\omega_1) > P(\omega_2)$

Decision Rule based on ONLY Prior

- $P(\omega = \omega_1)$, or $P(\omega_1)$ for prior next is the sea bass.
- $P(\omega = \omega_2)$, or $P(\omega_2)$ for prior next is the salmon.
- $P(\omega_1) + P(\omega_2) = 1$: either ω_1 or ω_2 must occur.
- A decision rule prescribes what actions to take based on observed data.
- Assume **only prior** available and **equal cost** for incorrect classifications.
  - Decide ω_1 if $P(\omega_1) > P(\omega_2)$
  - Otherwise, decide ω_2

Decision Rule based on ONLY Prior

- $P(\omega = \omega_1)$, or $P(\omega_1)$ for prior next is the sea bass.
- $P(\omega = \omega_2)$, or $P(\omega_2)$ for prior next is the salmon.
- $P(\omega_1) + P(\omega_2) = 1$: either ω_1 or ω_2 must occur.
- A decision rule prescribes what actions to take based on observed data.
- Assume **only prior** available and **equal cost** for incorrect classifications.
  - Decide ω_1 if $P(\omega_1) > P(\omega_2)$
  - Otherwise, decide ω_2

Limitation: Always choose the same. If the prior is uniform (e.g., $P(\omega_1) = P(\omega_2) = 0.5$), such rule behaves not well.

Class Conditional Density

- Use class-conditional information could improve accuracy.
- A **feature** is an observable variable, e.g., lightness, length, width, etc.
- Class Conditional Density: probability density function for x, the feature, given the state of nature is ω, i.e.,$p(x|\omega)$
- E.g., $p(x|\omega_1), p(x|\omega_2)$ describe the difference in lightness between populations of sea bass and salmon

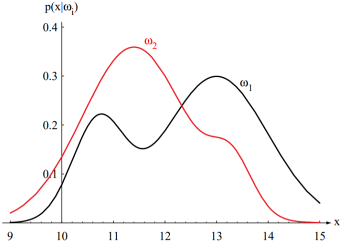

Figure: Class conditional probability[DHS book chapter 2]

Posterior Probability

- If know prior $P(\omega)$ and conditional density $p(x|\omega)$, as well as observed feature value x (e.g., lightness of the fish), how does that affect our decision?

Posterior Probability

- If know prior $P(\omega)$ and conditional density $p(x|\omega)$, as well as observed feature value x (e.g., lightness of the fish), how does that affect our decision?
- Posterior probability: the probability of a certain state of nature ω given our observables feature x: $P(\omega|x)$

Posterior Probability

- If know prior $P(\omega)$ and conditional density $p(x|\omega)$, as well as observed feature value x (e.g., lightness of the fish), how does that affect our decision?
- Posterior probability: the probability of a certain state of nature ω given our observables feature x: $P(\omega|x)$
- Bayes rule:

$$
\begin{aligned}
P(\omega_i|x) &= \frac{p(x|\omega_i)P(\omega_i)}{p(x)} \\
p(x) &= \sum_{i=1}^{2} p(x|\omega_i)P(\omega_i) \\
posterior &= \frac{likelihood \times prior}{evidence}
\end{aligned}
$$

Posterior Probability

- Posterior is determined by prior and likelihood.
- Example: when $P(\omega_1) = \frac{2}{3}$, $P(\omega_2) = \frac{1}{3}$

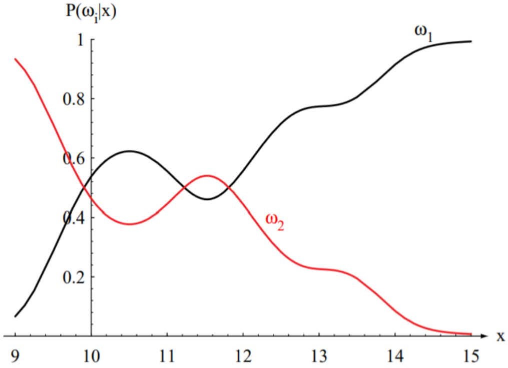

Figure: Posterior probability[DHS book chapter 2]

Decision Rule based on Posterior

- Given observation x, the decision is based on posterior probability.
  - Decide ω_1, if $P(\omega_1|x) > P(\omega_2|x)$
  - Decide ω_2, if $P(\omega_2|x) > P(\omega_1|x)$

Decision Rule based on Posterior

- Given observation x, the decision is based on posterior probability.
  - Decide ω_1, if $P(\omega_1|x) > P(\omega_2|x)$
  - Decide ω_2, if $P(\omega_2|x) > P(\omega_1|x)$
- Probability of error: for two class scenario, whenever we observe a particular x,

$$P(error|x) = \begin{cases} P(\omega_1|x), \text{if decide } \omega_2 \\ P(\omega_2|x), \text{if decide } \omega_1 \end{cases}$$

Minimize the Probability of Error

- Minimize the probability of error.
- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; otherwise decide ω_2.
- $P(error|x) = \min[P(\omega_1|x), P(\omega_2|x)]$
- Also minimize the average probability of error:

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error|x)p(x)dx$$

Bayesian Decision Rule

- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; otherwise decide ω_2.
- (Equivalent):
  Decide ω_1, if $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$

Bayesian Decision Rule

- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; otherwise decide ω_2.
- (Equivalent):
  Decide ω_1, if $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$
  - *evidence* $p(x)$: unimportant for making a decision.

Bayesian Decision Rule

- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; otherwise decide ω_2.
- (Equivalent):
  Decide ω_1, if $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$
  - *evidence* $p(x)$: unimportant for making a decision.
  - If for some x, we have $p(x|\omega_1) = p(x|\omega_2) \rightarrow$ decision rely on prior.

Bayesian Decision Rule

- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; otherwise decide ω_2.
- (Equivalent):
  Decide ω_1, if $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$
  - *evidence* $p(x)$: unimportant for making a decision.
  - If for some x, we have $p(x|\omega_1) = p(x|\omega_2) \rightarrow$ decision rely on prior.
  - If have uniform prior$\rightarrow$ decision rely on likelihood.

Bayesian Decision Rule

- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; otherwise decide ω_2.
- (Equivalent):
  Decide ω_1, if $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$
  - *evidence* $p(x)$: unimportant for making a decision.
  - If for some x, we have $p(x|\omega_1) = p(x|\omega_2) \rightarrow$ decision rely on prior.
  - If have uniform prior$\rightarrow$ decision rely on likelihood.
- Assumption: equal cost for each decision.

Bayesian Decision Rule

- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; otherwise decide ω_2.
- (Equivalent):
  Decide ω_1, if $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$
  - *evidence* $p(x)$: unimportant for making a decision.
  - If for some x, we have $p(x|\omega_1) = p(x|\omega_2) \rightarrow$ decision rely on prior.
  - If have uniform prior$\rightarrow$ decision rely on likelihood.
- Assumption: equal cost for each decision.
- Summary: Given both prior and likelihoods, Bayesian decision rule combines them (through posterior probability) for decision making which achieves minimum probability of error.

Bayesian Decision Theory

Bayesian Decision Theory-Continuous Feature

Generalize the previous fish example in several ways:

- allow the use of more than one feature. (length, weight etc)
- allow more than two states of nature. (tilapia, sardine etc)
- allow actions other than deciding the state of nature. (Not make a decision)
- introduce loss function more general than the probability of error. (some classification mistakes are more costly than others)

Notation

- feature vector $\mathbf{x} = (x_1, x_2, ..., x_d) \in R^d$: allow use of more than one feature.
- $\omega_1, \omega_2, ..., \omega_c$: finite set of c states of nature, i.e., categories.
- $\alpha_1, \alpha_2, ..., \alpha_a$: finite set of a possible actions.
- $\lambda(\alpha_i|\omega_i)$: loss function, describes the loss incurred for taking action α_i when state of nature is ω_i.
- $P(\omega_i)$: prior probability that state of nature is ω_i.
- $p(\mathbf{x}|\omega_i)$: state conditional probability for $\mathbf{x}$.

Posterior Probability

Bayes formula:

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}$$

The evidence $p(\mathbf{x})$:

$$p(\mathbf{x}) = \sum_{i=1}^{c} p(\mathbf{x}|\omega_i)P(\omega_i)$$

Conditional Risk

- Observe $\mathbf{x}$, take action α_i, if true state of nature $\omega_j \rightarrow$ loss $\lambda(\alpha_i|\omega_j)$.
- The expected loss, or conditional risk, of taking action α_i is (on board):

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

- For given observation $\mathbf{x}$, selecting the action that minimizes the conditional risk.

Overall Risk

- **Decision rule**: function $\alpha(\mathbf{x})$: $R^d \rightarrow \{\alpha_1, ..., \alpha_a\}$, indicate which action to take for every possible observation $\mathbf{x}$.
- The overall risk: expected loss associated with a given decision rule $\alpha(\mathbf{x})$ considering all possible observations:

$$R = \int R(\alpha(\mathbf{x})|\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

- Choose $\alpha(\mathbf{x})$ that minimizes the overall risk.

Bayesian Decision Rule

Compute conditional risk for all possible actions:

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

Select action α_i that has minimum conditional risk:

$$\alpha^\star = \arg\min_{\alpha_i} R(\alpha_i|\mathbf{x})$$

- Bayesian decision rule minimizes the overall risk. (**Minimum Risk Decision**)
- The minimum overall risk $R^\star$ is called **Bayes risk**, best performance we can get.

Two Category Classification

Two Class Classification

- α_1: deciding that the true state of nature is ω_1.
  α_2: deciding that the true state of nature is ω_2.
- $\lambda(\alpha_i|\omega_j)$: loss incurred for deciding ω_i when the true state of nature is ω_j, denote as λ_{ij}.

Two Class Classification

- α_1: deciding that the true state of nature is ω_1.
  α_2: deciding that the true state of nature is ω_2.
- $\lambda(\alpha_i|\omega_j)$: loss incurred for deciding ω_i when the true state of nature is ω_j, denote as λ_{ij}.
- Recall $R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x})$.
  - $R(\alpha_1|\mathbf{x}) = \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x})$
  - $R(\alpha_2|\mathbf{x}) = \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x})$

Two Class Classification

- α_1: deciding that the true state of nature is ω_1.
  α_2: deciding that the true state of nature is ω_2.
- $\lambda(\alpha_i|\omega_j)$: loss incurred for deciding ω_i when the true state of nature is ω_j, denote as λ_{ij}.
- Recall $R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x})$.
  - $R(\alpha_1|\mathbf{x}) = \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x})$
  - $R(\alpha_2|\mathbf{x}) = \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x})$
- Decision Rule: decide ω_1 if $R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x})$

Two Class Classification

- α_1: deciding that the true state of nature is ω_1.
  α_2: deciding that the true state of nature is ω_2.
- $\lambda(\alpha_i|\omega_j)$: loss incurred for deciding ω_i when the true state of nature is ω_j, denote as λ_{ij}.
- Recall $R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x})$.
  - $R(\alpha_1|\mathbf{x}) = \lambda_{11} P(\omega_1|\mathbf{x}) + \lambda_{12} P(\omega_2|\mathbf{x})$
  - $R(\alpha_2|\mathbf{x}) = \lambda_{21} P(\omega_1|\mathbf{x}) + \lambda_{22} P(\omega_2|\mathbf{x})$
- Decision Rule: decide ω_1 if $R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x})$
- In terms of posterior:

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

Likelihood Ratio

- In terms of prior and likelihood:

$$(\lambda_{21} - \lambda_{11})P(\omega_1|\mathbf{x}) \quad > \quad (\lambda_{12} - \lambda_{22})P(\omega_2|\mathbf{x})$$

Likelihood Ratio

- In terms of prior and likelihood:

$$
\begin{aligned}
(\lambda_{21} - \lambda_{11})P(\omega_1|\mathbf{x}) &> (\lambda_{12} - \lambda_{22})P(\omega_2|\mathbf{x}) \\
(\lambda_{21} - \lambda_{11})P(\mathbf{x}|\omega_1)P(\omega_1) &> (\lambda_{12} - \lambda_{22})P(\mathbf{x}|\omega_2)P(\omega_2)
\end{aligned}
$$

- Assume: $\lambda_{21} > \lambda_{11}$ and $\lambda_{12} > \lambda_{22}$ (loss incurred for making an mistake is greater than loss incurred for being correct)

Likelihood Ratio

- In terms of prior and likelihood:

$$
\begin{aligned}
(\lambda_{21} - \lambda_{11})P(\omega_1|\mathbf{x}) &> (\lambda_{12} - \lambda_{22})P(\omega_2|\mathbf{x}) \\
(\lambda_{21} - \lambda_{11})P(\mathbf{x}|\omega_1)P(\omega_1) &> (\lambda_{12} - \lambda_{22})P(\mathbf{x}|\omega_2)P(\omega_2)
\end{aligned}
$$

- Assume: $\lambda_{21} > \lambda_{11}$ and $\lambda_{12} > \lambda_{22}$ (loss incurred for making an mistake is greater than loss incurred for being correct)
- **Likelihood ratio test**: decide ω_1 if

$$
\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \underbrace{\frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}}_{\text{independent of } \mathbf{x}}
$$

Example

- $p(\mathbf{x}|\omega_1) = N(4, 1)$, $p(\mathbf{x}|\omega_2) = N(10, 1)$
- $P(\omega_1) = \frac{1}{3}$
- $\lambda = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$
- Decision rule?

Minimum Error Rate

Minimum-Error-Rate Classification

- Actions are decisions on classes
  - If action α_i is taken and true state of nature is ω_j, then decision correct if $i = j$, and in error if $i \neq j$.
- Choose the decision rule that minimizes the probability of error, i.e., *error rate*.

Zero-One Loss

Zero-one Loss:

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0, \text{if } i = j \\ 1, \text{if } i \neq j \end{cases}$$

- NO cost for correct decision.
- SAME unit cost for any errors.

Risk

Conditional risk:

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x}) \\ &= \sum_{i \neq j} P(\omega_j|\mathbf{x}) \end{aligned}$$

Risk

Conditional risk:

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x}) \\ &= \sum_{i \neq j} P(\omega_j|\mathbf{x}) \\ &= 1 - P(\omega_i|\mathbf{x}) \end{aligned}$$

$P(\omega_i|\mathbf{x})$: posterior probability that action α_i is correct given observation $\mathbf{x}$.

Bayesian Decision Rule

- **Minimum Risk Decision**: choose the action that minimize the conditional risk. $(R(\alpha_i|\mathbf{x}) = 1 - P(\omega_i|\mathbf{x}))$

$$\min R(\alpha_i|\mathbf{x}) \equiv \max P(\omega_i|\mathbf{x})$$

- Decide ω_i if $P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x})$ for $j \neq i$.
- Above rule: minimize probability of error, minimize error rate, minimize the risk etc

Likelihood Ratio

- **Likelihood Ratio Test**:

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

- Let $\theta_\lambda = \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$, then
  Decide ω_1 if $\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \theta_\lambda$
- For zero-one loss: $\lambda = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\theta_\lambda = \frac{P(\omega_2)}{P(\omega_1)} = \theta_a$
- Penalize more on misclassifying ω_2 to ω_1, e.g., $\lambda = \begin{bmatrix} 0 & 5 \\ 1 & 0 \end{bmatrix}$,
  $\theta_\lambda = \frac{5P(\omega_2)}{P(\omega_1)} = \theta_b$

Likelihood Ratio for fish example

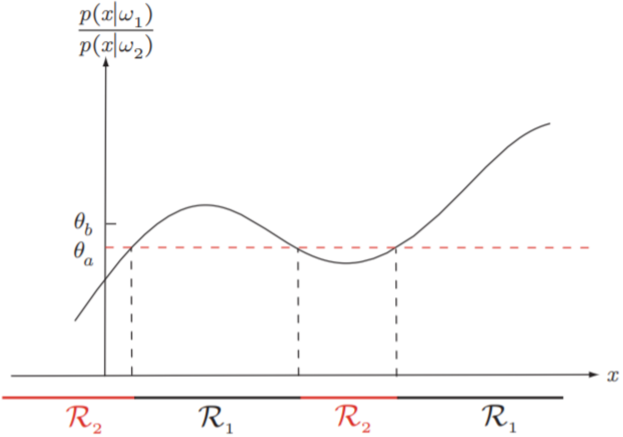

Figure: Likelihood Ratio. If use zero-one loss, the decision boundary is determined by threshold θ_a.[DHS book chapter 2]

Discriminant Functions

Discriminant Function

- Discriminant functions: useful way to represent pattern classifier.
- $g_i(\mathbf{x})$: discriminant function for i-th class.
- The classifier is said to assign an observation (or feature vector) $\mathbf{x}$ to class ω_i if:

$$g_i(\mathbf{x}) > g_j(\mathbf{x}), \text{for } j \neq i$$

- Decide ω_i that have **largest** discriminant.

Network Representation of Classifier

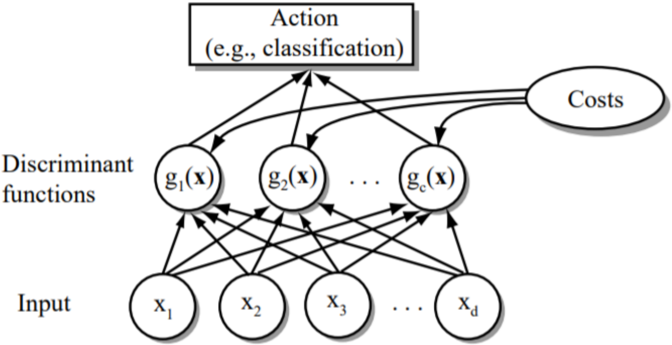

Figure: Classifier which includes d inputs and c discriminant function $g_i(\mathbf{x})$ [DHS book chapter 2]

Bayesian Classifier

Bayesian classifier can be naturally represented using discriminants:

Bayesian Classifier

Bayesian classifier can be naturally represented using discriminants:

- General risk: $g_i(\mathbf{x}) = -R(\alpha_i|\mathbf{x})$
  Maximum discriminant function is equivalent to **minimum** conditional risk.

Bayesian Classifier

Bayesian classifier can be naturally represented using discriminants:

- General risk: $g_i(\mathbf{x}) = -R(\alpha_i|\mathbf{x})$
  Maximum discriminant function is equivalent to **minimum** conditional risk.
- Zero-one loss: $g_i(\mathbf{x}) = P(\omega_i|\mathbf{x})$.
  Maximum discriminant function is equivalent to **maximum** posterior probability.

Choice of Discriminant Function

- The choice of discriminant function is NOT unique.
  - Multiply by some positive constant
  - Shift by some constant
  - Use monotone increasing function $f(.)$ on $g_i(\mathbf{x})$
- Particularly:

$$g_i(\mathbf{x}) = P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} p(\mathbf{x}|\omega_i)P(\omega_i)}$$

$$g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i)$$

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$$

Two Category Case

Usually define a single discriminant function

$$g(\mathbf{x}) \equiv g_1(\mathbf{x}) - g_2(\mathbf{x})$$

For minimum-error-rate (i.e., with zero-one loss), followings are convenient:

$$g(\mathbf{x}) = P(\omega_1|\mathbf{x}) - P(\omega_2|\mathbf{x})$$

$$g(\mathbf{x}) = \ln \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)}$$

Decision rule: decide ω_1 if $g(\mathbf{x}) > 0$, otherwise decide ω_2.

Decision Region

- Discriminant functions of various forms, same decision rules.
- Decision rule divides the feature space $\mathbf{x} \in R^d$ into c **decision regions**: $R_1, ..., R_c$, separated by *decision boundaries*.

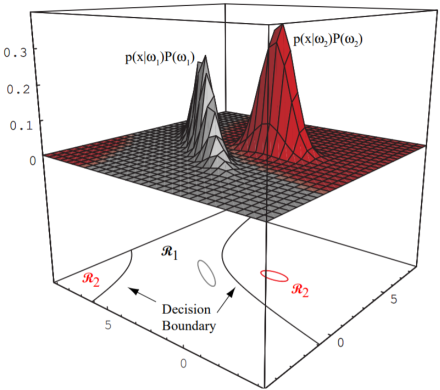

Figure: [DHS book chapter 2]

Three Approaches for Decision Problem

Three Approaches

In general, we have three distinct approaches for decision problem which are (in increasing order of complexity):

Three Approaches

In general, we have three distinct approaches for decision problem which are (in increasing order of complexity):

- **Discriminant function**: find a function $g(\mathbf{x})$ which maps each input $\mathbf{x}$ directly onto a class label.

Three Approaches

In general, we have three distinct approaches for decision problem which are (in increasing order of complexity):

- **Discriminant function**: find a function $g(\mathbf{x})$ which maps each input $\mathbf{x}$ directly onto a class label.
- **Discriminative models**: approaches that model the posterior probabilities directly (i.e., $p(\omega_k|\mathbf{x})$).

Three Approaches

In general, we have three distinct approaches for decision problem which are (in increasing order of complexity):

- **Discriminant function**: find a function $g(\mathbf{x})$ which maps each input $\mathbf{x}$ directly onto a class label.
- **Discriminative models**: approaches that model the posterior probabilities directly (i.e., $p(\omega_k|\mathbf{x})$).
- **Generative models**: approaches that model the joint distribution $p(\omega_k, \mathbf{x})$.
  - Specifically, determining the class-conditional densities $p(\mathbf{x}|\omega_k)$ and prior class probabilities $p(\omega_k)$ for each class ω_k individually. Then use Bayes' theorem to find posterior $p(\omega_k|\mathbf{x})$.