Johns Hopkins
Whiting School of Engineering

Applied and Computational Mathematics

Data Mining
625.740
Pattern Classification

Mike Weisman

email: data.mining.625.740@gmail.com

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Applied and Computational Mathematics

Data Mining

625.740

Pattern Classification

States and Actions

Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ be the finite set of m states of nature and let $A = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be the finite set of k possible actions.

Let $\lambda(\alpha_a|\omega_j)$ be the loss incurred for taking action α_a when the state is ω_j.

Let the feature vector $\mathbf{x}$ be an n-component vector valued random variable, and let $p(\mathbf{x}|\omega_j)$ be the class-conditioned probability density function for $\mathbf{x}$. Let $P(\omega_j)$ be the *a priori* probabilities for each of the states ω_j.

Then each *a posteriori* probability $p(\omega_j|x)$ can be computed by Bayes Rule:

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

$$\text{where } p(\mathbf{x}) = \sum_{q=1}^{m} p(\mathbf{x}|\omega_q)P(\omega_q).$$

Conditional Risk

Suppose we observe a particular $\mathbf{x}$ and we contemplate taking action α_a. If the true class is ω_j, we will incur the loss $\lambda(\alpha_a|\omega_j)$. Since $p(\omega_j|\mathbf{x})$ is the probability that the true class is ω_j, the expected loss associated with taking action α_a is

$$R(\alpha_a|\mathbf{x}) = \sum_{j=1}^{m} \lambda(\alpha_a|\omega_j) p(\omega_j|\mathbf{x}).$$

$R(\alpha_a|\mathbf{x})$ is known as the conditional risk.

Decision Rule

A decision rule is a function $\alpha(\mathbf{x})$ that tells us which action to take for every possilbe observation.

$\forall\mathbf{x}$, the decision function $\alpha(\mathbf{x})$ assumes one of the k values $\alpha_1, \alpha_2, \ldots, \alpha_k$.

The overall risk R is the expected loss associated with a given decision rule:

$$R = \int \ldots \int_{\mathbb{R}^n} R(\alpha(\mathbf{x})|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}.$$

Bayes Risk

To minimize overall risk R, choose $\alpha(\mathbf{x})$ so $R(\alpha(\mathbf{x})|\mathbf{x})$ is minimum for every $\mathbf{x}$.

Bayes Decision Rule:

$$\alpha_a^* = \arg\min_{\alpha_a} R(\alpha_a|\mathbf{x}) = \arg\min_{\alpha_a} \sum_{j=1}^{m} \lambda(\alpha_a|\omega_j)p(\omega_j|\mathbf{x}).$$

The resulting overall risk is called the Bayes risk and is the best performance that can be achieved.

Two Category Classification

Action α_1 corresponds to deciding class ω_1 and action α_2 corresponds to deciding class ω_2. Let $\lambda_{aj} = \lambda(\alpha_a|\omega_j)$.
The conditional risk can now be written

$$R(\alpha_1|\mathbf{x}) = \lambda_{11}p(\omega_1|\mathbf{x}) + \lambda_{12}p(\omega_2|\mathbf{x})$$
$$R(\alpha_2|\mathbf{x}) = \lambda_{21}p(\omega_1|\mathbf{x}) + \lambda_{22}p(\omega_2|\mathbf{x})$$

Deciding ω_1 if $R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x})$ yields

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

Equivalently

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

Likelihood Ratio

Assuming that $\lambda_{21} > \lambda_{11}$

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

The conditional probability $p(\mathbf{x}|\omega_j)$ is called the likelihood of ω_j with respect to $\mathbf{x}$ and $\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)}$ is called the likelihood ratio.

Bayes Decision Rule is to decide ω_1 if the likelihood ratio is greater than a threshold that is independent of $\mathbf{x}$.

Minimum Error Rate Classification

Assume m states of nature $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ and let α_j be the decision that we are in state ω_j. If action α_a is taken, our decision is correct if $a = j$ and in error if $a \neq j$. We seek a decision rule that minimizes the error rate = average probabilty of error.
Consider the loss function

$$\lambda(\alpha_a|\omega_j) = \begin{cases} 0, a = j \\ 1, a \neq j \end{cases} \quad (a = 1, \ldots, m; \quad j = 1, \ldots, m)$$

The conditional risk is

$$\begin{aligned} R(\alpha_a|\mathbf{x}) &= \sum_{j=1}^{m} \lambda(\alpha_a|\omega_j) p(\omega_j|\mathbf{x}) \\ &= \sum_{j \neq a} p(\omega_j|\mathbf{x}) \\ &= 1 - p(\omega_a|\mathbf{x}) \end{aligned}$$

Minimimum Error Rate Classification

The quantity $p(\omega_a|\mathbf{x})$ is the conditional probability that action α_a is correct.
Bayes decision rule: Minimize overall risk by minimizing conditional risk.

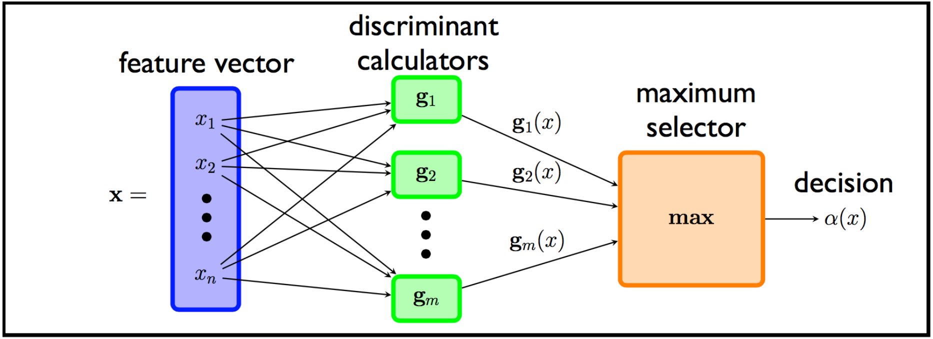
Decide ω_a if $p(\omega_a|\mathbf{x}) > p(\omega_j|\mathbf{x}) \quad \forall a \neq j$.

This gives the minimum error rate.

Pattern Classifier

We define a set of discriminant functions $g_j(\mathbf{x})$, $\quad j = 1, \ldots, m,$ $\quad$ and we assign feature vector $\mathbf{x}$ to class ω_j when

$$g_j(\mathbf{x}) > g_a(\mathbf{x}), \quad \forall a \neq j$$



General case: $g_j(\mathbf{x}) = -R(\alpha_j|\mathbf{x})$

Minimum error rate classifier: $g_j(\mathbf{x}) = p(\omega_j|\mathbf{x})$ $\quad$ (max.*a posteriori* probability)

Pattern Classifier

If f is a monotonically increasing function, we can replace $g_j(\mathbf{x})$ by $f(g_j(\mathbf{x}))$.

For minimum-error-rate classification, these all give identical classification results:

$$g_j(\mathbf{x}) = p(\omega_j|\mathbf{x});$$
$$g_j(\mathbf{x}) = p(\mathbf{x}|\omega_j)P(\omega_j);$$
$$g_j(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{\sum_{a=1}^{m} p(\mathbf{x}|\omega_a)P(\omega_a)};$$
$$g_j(\mathbf{x}) = \log p(\mathbf{x}|\omega_j) + \log P(\omega_j);$$

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Applied and Computational Mathematics

Data Mining

625.740

Pattern Classification