JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Applied and Computational Mathematics

Data Mining
625.740
Bayes Decision Theory

Mike Weisman

email: `data.mining.625.740@gmail.com`

JOHNS HOPKINS
WHITING SCHOOL *of* ENGINEERING

Applied and Computational Mathematics

Data Mining

625.740

Bayes Decision Theory

Bayes Decision Theory

The sushi problem: Suppose there are two types (or classes) of sushi

$$\omega = \begin{cases} \omega_1, & \text{for salmon} \\ \omega_2, & \text{for tuna} \end{cases}$$

Assuming we know the *a priori* probabilities $P(\omega_1)$ & $P(\omega_2)$ of each type, we wish to predict what the next order of sushi will be!

Bayes Decision Theory

The sushi problem: Suppose there are two types (or classes) of sushi

$$\omega = \begin{cases} \omega_1, \text{ for salmon} \\ \omega_2, \text{ for tuna} \end{cases}$$

Assuming we know the *a priori* probabilities $P(\omega_1)$ & $P(\omega_2)$ of each type, we wish to predict what the next order of sushi will be!

Decision Rule

$$\text{Decide } \begin{cases} \omega_1, \text{ if } P(\omega_1) > P(\omega_2) \\ \omega_2, \text{ otherwise} \end{cases}$$

Features

Given feature vectors $\mathbf{x}$, suppose we know the class-conditioned probability functions $p(\mathbf{x}|\omega_j), \quad j = 1, 2$
e.g.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \text{red} \\ \text{green} \\ \text{blue} \end{pmatrix} = \textbf{color}$$

Bayes Rule

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

$$\text{where } p(\mathbf{x}) = \sum_{k=1}^{2} p(\mathbf{x}|\omega_k)P(\omega_k)$$

Probability of Error

$$p(\text{error}|\mathbf{x}) = \begin{cases} p(\omega_1|\mathbf{x}), & \text{if we decide } \omega_2 \\ p(\omega_2|\mathbf{x}), & \text{if we decide } \omega_1 \end{cases}$$

Probability of Error

$$p(\text{error}|\mathbf{x}) = \begin{cases} p(\omega_1|\mathbf{x}), & \text{if we decide } \omega_2 \\ p(\omega_2|\mathbf{x}), & \text{if we decide } \omega_1 \end{cases}$$

For all $\mathbf{x}$ we can minimize $p(\text{error}|\mathbf{x})$ if we

$$\text{Decide} \begin{cases} \omega_1, & \text{when } p(\omega_1|\mathbf{x}) > p(\omega_2|\mathbf{x}) \\ \omega_2, & \text{otherwise} \end{cases} \qquad (*)$$

Bayes Decision Rule

$$p(\text{error}|\mathbf{x}) = \begin{cases} p(\omega_1|\mathbf{x}), & \text{if we decide } \omega_2 \\ p(\omega_2|\mathbf{x}), & \text{if we decide } \omega_1 \end{cases}$$

For all $\mathbf{x}$ we can minimize $p(\text{error}|\mathbf{x})$ if we

$$\text{Decide } \begin{cases} \omega_1, & \text{when } p(\omega_1|\mathbf{x}) > p(\omega_2|\mathbf{x}) \\ \omega_2, & \text{otherwise} \end{cases} \qquad (*)$$

The average probability of error is

$$P_e = \iiint_{\mathbb{R}^3} p(\text{error}, \mathbf{x})\, d\mathbf{x} = \iiint_{\mathbb{R}^3} p(\text{error}|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}$$

Since equation $(*)$ minimizes $p(\text{error}|\mathbf{x})$ for all $\mathbf{x}$, it minimizes P_e.
The decision given in equation $(*)$ is known as the Bayes Decision Rule.

Bayes Decision Rule

Applying Bayes Rule and disregarding the "scale factor" $p(\mathbf{x})$, The Bayes Decision Rule $(*)$ becomes

$$\text{Decide} \begin{cases} \omega_1, \text{ if } p(\mathbf{x}|\omega_1)P(\omega_1) > p(\mathbf{x}|\omega_2)P(\omega_2) \\ \omega_2, \text{ otherwise} \end{cases} \qquad (**)$$

If $\exists \mathbf{x} \ni p(\mathbf{x}|\omega_1) = p(\mathbf{x}|\omega_2)$, then the decision hinges only on the *a priori* probabilities.

If $P(\omega_1) = P(\omega_2)$, then the decision is based entirely on $p(\mathbf{x}|\omega_j)$, the likelihoods of ω_j with respect to $\mathbf{x}$.

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Applied and Computational Mathematics

Data Mining

625.740

Bayes Decision Theory