Multivariate Gaussian Distribution Model

Robert B. Fisher
School of Informatics
University of Edinburgh

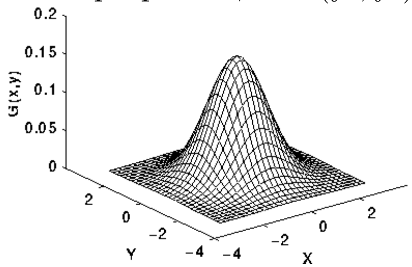
$prob(\vec{x}|c)$? Gaussian Distribution

Data is feature vector $\vec{x} = (f_1, f_2, \ldots, f_n)'$.

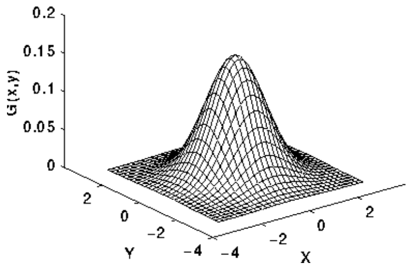
Expect variation in property values over the class,
perhaps not independent between variables.

Commonly [?] probability distribution of $\vec{x}$
is Multivariate Gaussian Distribution

For 2 properties, $\vec{x} = (f_1, f_2)'$ we have:



2D Gaussian Distribution



Characterised by mean $(m_1, m_2)^T$ and covariance matrix

$$\begin{bmatrix} (\sigma_1)^2 & \rho_{ij}\sigma_1\sigma_2 \\ \rho_{ij}\sigma_1\sigma_2 & (\sigma_2)^2 \end{bmatrix}$$

σ_i - standard [?] of i^{th} property

ρ_{ij} - cross-correlation coefficient between i and j

Multivariate Gaussian Distribution

For each class c we need:

- Mean vector $\vec{m}_c$ of dimension n - the [?] value of the n properties for class c
- Covariance matrix $\mathcal{A}_c$ - the $n \times n$ matrix of joint variation between each pair of properties.

Then, the probability of observing feature vector $\vec{x}$ given class c is:

$$p(\vec{x}|c) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{|\mathcal{A}_c|^{\frac{1}{2}}} e^{-\frac{1}{2}[(\vec{x}-\vec{m}_c)^T \mathcal{A}_c^{-1} (\vec{x}-\vec{m}_c)]}$$

Estimating the Distribution Parameters: the Class Model

Given $k > n$ training instances $\{\vec{x}_i\}$ of class c

Estimated Mean vector:

$$\vec{m}_c = \frac{1}{k}\sum_{i=1}^{k}\vec{x}_i$$

Estimated [?] matrix:

$$\mathcal{A}_c = \frac{1}{k-1}\sum_{i=1}^{k}(\vec{x}_i - \vec{m}_c)(\vec{x}_i - \vec{m}_c)^T$$

Estimate $p(c)$ from class distribution of training samples

 Slide based on: Bob Fisher

Probability Example

Two classes. *A priori* probabilities p(1) = 0.6, p(2) = 0.4

Cls 1: $\vec{m}_1 = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ $\mathcal{A}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ $\mathcal{A}_1^{-1} = \frac{1}{5}\begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$

Cls 2: $\vec{m}_2 = \begin{bmatrix} 4 \\ 13 \end{bmatrix}$ $\mathcal{A}_2 = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ $\mathcal{A}_2^{-1} = \frac{1}{5}\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$

$|\mathcal{A}_1| = |\mathcal{A}_2| = 5$

Test data $\vec{x} = \begin{bmatrix} area \\ perimeter \end{bmatrix} = \begin{bmatrix} 3 \\ 10 \end{bmatrix}$

Which is the most probable [?]?

 Slide based on: Bob Fisher

Example continued

Class 1 if $p(1|\vec{x}) > p(2|\vec{x})$

[?] rule:

$$p(1|\vec{x}) = \frac{p(\vec{x}|1)p(1)}{p(\vec{x})} > \frac{p(\vec{x}|2)p(2)}{p(\vec{x})} = p(2|\vec{x})$$

$$0.6 \cdot p(\vec{x}|1) > 0.4 \cdot p(\vec{x}|2)$$

$$p(\vec{x}|1) = \frac{1}{2\pi}\frac{1}{|\mathcal{A}_1|^{\frac{1}{2}}}e^{-\frac{1}{2}[(\vec{x}-\vec{m}_1)^T\mathcal{A}_1^{-1}(\vec{x}-\vec{m}_1)]}$$

 Slide based on: Bob Fisher

$$= \frac{1}{2\pi}\frac{1}{5^{\frac{1}{2}}}e^{-\frac{1}{2}\left[\left(\begin{bmatrix} 3 \\ 10 \end{bmatrix}-\begin{bmatrix} 2 \\ 6 \end{bmatrix}\right)^T \mathcal{A}_1^{-1}\left(\begin{bmatrix} 3 \\ 10 \end{bmatrix}-\begin{bmatrix} 2 \\ 6 \end{bmatrix}\right)\right]}$$

$$= \frac{1}{2\pi}\frac{1}{\sqrt{5}}e^{-\frac{1}{2}\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix}^T \frac{1}{5}\begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 4 \end{bmatrix}\right)}$$

$$p(\vec{x}|1) = \frac{1}{2\pi}\frac{1}{\sqrt{5}}e^{-\frac{27}{10}} = 4.78 \cdot 10^{-3}$$

 Slide based on: Bob Fisher

Similarly,

$$p(\vec{x}|2) = \frac{1}{2\pi}\frac{1}{\sqrt{5}}e^{-\frac{23}{10}} = 7.15 \cdot 10^{-3}$$

So, class 1 if

$$0.6p(\vec{x}|1) > 0.4p(\vec{x}|2)$$

$$2.87 \cdot 10^{-3} > 2.85 \cdot 10^{-3}$$

Thus most [?] (barely) to be class 1.

Lecture Overview

1. Multivariate Gaussian distribution
2. Estimating distribution parameters: mean and [?]