Bayesian Decision Theory

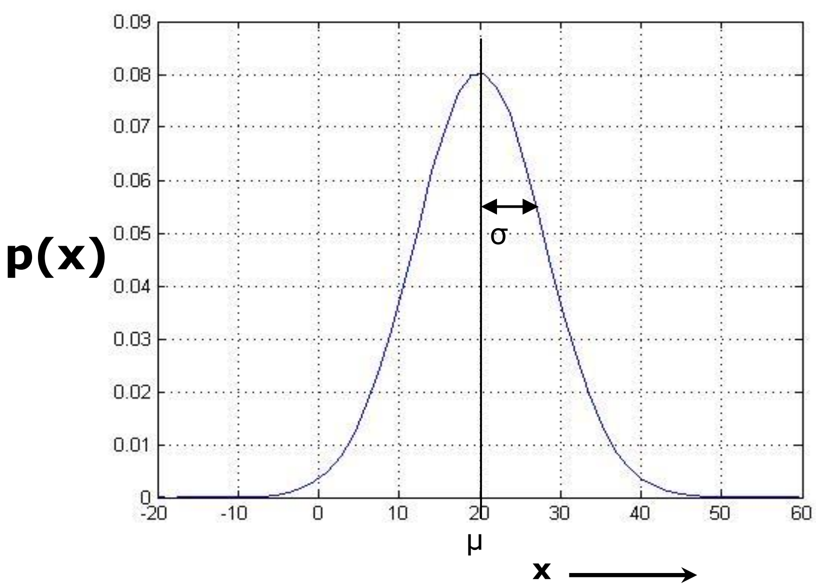
Dr. Deepak Ranjan Nayak
SVNIT Surat
Email: drnayak@ieee.org

Probability Distribution

- Data of a class is represented by a probability distribution
- For a class whose data is considered to be forming a single cluster, it can be represented by a normal or Gaussian distribution
- Univariate Gaussian distribution:

$$p(x) = N(x / \mu, \sigma)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



- μ is the mean
- σ^2 is the variance

Gaussian Distribution

Gaussian pdf is extensively used in pattern recognition and machine learning because of

- Mathematical tractability
- Central limit theorem

- Normal distribution makes math easy-things like calculating moments, correlations between variables, and other calculations that are domain specific.
- In simple terms, if you have many independent variables that may be generated by all kinds of distributions, assuming that no thing too crazy happens, the aggregate of those variables will tend toward a normal distribution.

Multivariate Gaussian Distribution

- **Data in *d*-dimensional space**

$$p(\mathbf{x}) = N(\mathbf{x}/\mu, \mathrm{C}) = \frac{1}{(2\pi)^{d/2}|C|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^{t}\mathrm{C}^{-1}(\mathbf{x}-\mu)}$$

- **μ is the mean vector**
- **C is the covariance matrix**

$$d = 2$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} E[x_1] \\ E[x_2] \end{bmatrix}$$

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

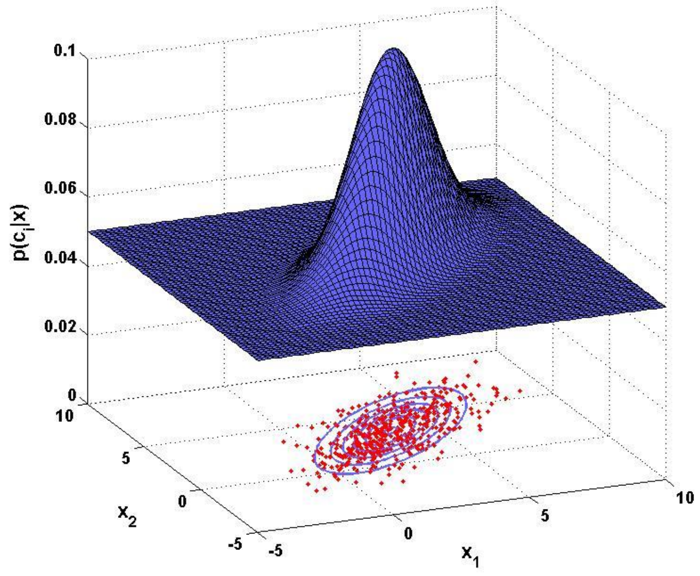
$$C = \begin{bmatrix} E\left[(x_1-\mu_1)^2\right] & E[(x_1-\mu_1)(x_2-\mu_2)] \\ E[(x_2-\mu_2)(x_1-\mu_1)] & E\left[(x_2-\mu_2)^2\right] \end{bmatrix}$$

Multivariate Gaussian Distribution

Data in *d*-dimensional space

- Size of mean vector μ : $\mathbf{1\times d}$
- Size of C the covariance matrix: $\boldsymbol{d\times d}$

Bivariate Gaussian distribution:



Question:

Compute the value of a Gaussian pdf, $N(\mu, C)$, at $x1 = [0.2, 1.3]^T$ and $x1 = [2.2, -1.3]^T$, where

$\mu = [0\ 1]^T$ **and** $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Solution:

Here, $d = 2$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|C|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^t C^{-1}(\mathbf{x}-\mu)}$$

$$p(\mathbf{x}1) = 0.1491$$

$$p(\mathbf{x}2) = 0.001$$

Bayes Classifier

- **Two-class pattern classification using Bayes classifier**
- **Data of a class is represented by its class-conditional probability density function (likelihood function)**
  - **Class 1:** $\mathrm{p}(\mathrm{x} \mid c_1)$
  - **Class 2:** $\mathrm{p}(\mathrm{x} \mid c_2)$
- **The *a priori* knowledge about the occurrence of classes is represented by the *prior* probabilities,** $\mathrm{P}(c_1)$ **and** $\mathrm{P}(c_2)$
- **The *a posteriori* probabilities are computed using the Bayes rule:**

$$\mathrm{P}(c_1 \mid \mathbf{x}) = \frac{\mathrm{p}(\mathbf{x} \mid c_1)\, \mathrm{P}(c_1)}{\mathrm{p}(\mathbf{x})}$$

$$\mathrm{P}(c_2 \mid \mathbf{x}) = \frac{\mathrm{p}(\mathbf{x} \mid c_2)\, \mathrm{P}(c_2)}{\mathrm{p}(\mathbf{x})}$$

- **Probability of occurrence of x,** $\mathrm{p}(\mathrm{x})$**, is the normalization factor**

Bayes Classifier (contd.)

- **For a test pattern, the class with the higher *a posteriori* probability is assigned using the following Bayes decision rule:**

$$\text{If } \mathrm{P}(c_1 \mid \mathbf{x}) > \mathrm{P}(c_2 \mid \mathbf{x}), \text{ assign } c_1 \text{ to } \mathbf{x};$$

$$\text{otherwise assign } c_2 \text{ to } \mathbf{x}$$

- **The above method can be extended for multi-class pattern classification.**

- **Bayes classifier is shown to give the minimum classification error when the probability distributions of classes are known.**

Bayes Classifier (contd.)

Generalized Bayes decision theory

- x is assigned to c_i

$$P(c_i|x) > P\left(c_j\middle|x\right), \qquad \forall j \neq i$$

- As $\mathrm{p}(x)$ is positive and the same for all classes

$$p(x|c_i)P(c_i) > p\left(x\middle|c_j\right)P\left(c_j\right), \qquad \forall j \neq i$$

Question:

Consider a 2-class classification task in the 2-dimensional space, where the data in both classes, *ω1*, *ω2*, are distributed according to the Gaussian distributions *N(*μ1,*C*1*)* and *N(*μ2,*C*2*)*

$$\mu 1 = [1, 1]^T,\ \mu 2 = [3, 3]^T,\ C1 = C2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Assuming that P(ω1) = P(ω2) = 1/2, classify x = [1.8, 1.8]^T into ω1 or ω2.

Solution:

$$P(\omega 1 \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega 1)\, P(\omega 1)}{p(\mathbf{x})}$$

$$P(\omega 2 \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega 2)\, P(\omega 2)}{p(\mathbf{x})}$$

$P(\omega 1 \mid x) = 0.042$

$P(\omega 2 \mid x) = 0.0189$

x is classified to ω1

Minimum Distance Classifier

Euclidean Distance Classifier

The optimal Bayesian classifier is significantly simplified under the following assumptions:

- The classes are equiprobable.
- The data in all classes follow Gaussian distributions.
- The covariance matrix is the same for all classes.
- The covariance matrix is diagonal and all elements across the diagonal are equal.

  That is, $C = \sigma^2 I$, I is the identity matrix.

x is assigned to c_i, if

$$\sqrt{(\boldsymbol{x}-\boldsymbol{\mu}_i)(\boldsymbol{x}-\boldsymbol{\mu}_i)^T} < \sqrt{(\boldsymbol{x}-\boldsymbol{\mu}_j)(\boldsymbol{x}-\boldsymbol{\mu}_j)^T} \quad \forall \boldsymbol{i} \neq \boldsymbol{j}$$

It assigns a pattern to the class whose mean is closest to it w.r.t the Euclidean norm.

Mahalanobis Distance Classifier

If one relaxes the assumptions required by the Euclidean classifier and removes the last one, the Bayesian classifier becomes equivalent to the minimum Mahalanobis distance classifier.

- ❖ The classes are equiprobable.
- ❖ The data in all classes follow Gaussian distributions.
- ❖ The covariance matrix is the same for all classes.

x is assigned to c_i, if

$$\sqrt{(\boldsymbol{x}-\boldsymbol{\mu}_i)\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\boldsymbol{T}}} < \sqrt{(\boldsymbol{x}-\boldsymbol{\mu}_j)\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)^{\boldsymbol{T}}} \ \ \forall \boldsymbol{i} \neq \boldsymbol{j}$$

Thank You