Chapter 1 Introduction

Pattern is a synthesis of measured values taken from a single sample of a limited part of the word.

Pattern Recognition is to attempt to determine the attributes of a sample type, that is, to attribute a certain sample to a class among multiple types.

The Model Space.

The Feature Space.

The Type Space.

Data collection -> Feature Extraction -> Feature Selection -> Classification Recognition

Machine learning algorithm is a kind of algorithm that automatically analyzes and obtains laws from data, and uses the laws to predict unknown data.

The objects in the training set in **supervised learning** are labeled by humans.

Unsupervised learning is characterized by only providing input samples. Common unsupervised learning algorithms include clustering, such as k-means algorithm.

Semi-supervised Learning

Ensemble learning is to obtain the final classification results by some combinations of the results of multiple classifiers.

Distribution function : $F(x) = P\{X < x\}$

Probability density function

Multivariate Normal Distribution

$$f(x) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

If a multivariate random variable $\mathbf{X}$ in n-dimensional space follows a normal distribution, also known as Gaussian distribution, with mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$, and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, then its probability density function (PDF) is:

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

where $|\cdot|$ denotes the determinant of a matrix.

Geometrically, the single Gaussian distribution model is approximately an ellipse in two-dimensional space and an ellipsoid in three-dimensional space. However, in many classification problems, a sample does not conform to the shape of an "ellipse". Therefore, we can divide a sample into different parts so that each part is similar to an "ellipse," which is then characterized by a Gaussian model. Consequently, the entire data can be described using a Gaussian mixture model.

The probability density function (PDF) of GMM can be expressed as:

$$f(\mathbf{x}) = \sum_{i=1}^{K} \pi_i \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

where π_i represents the weight or mixing coefficient for the i-th Gaussian component, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ represent the mean vector and covariance matrix of the i-th Gaussian component, respectively, and $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the PDF of a multivariate Gaussian distribution with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$.