1 : Implementing Bayes' Classifier

Author: Abijith J. Kamath   *Email:* abijithj@iisc.ac.in

Introduction

Let patterns be in $\mathbb{R}^d$ and consider the 2-class classification problem – given a pattern $\mathbf{x}$, classify the pattern into a class with label y. In Bayesian classification theory, the training data $\{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^n$ is viewed to be i.i.d samples of the random vector $(\mathbf{X}, Y)$. The training samples are used to model the joint distribution of the random vector $(\mathbf{X}, Y)$.

Let $\Pr(C_i) = p_i$ be the priors associated with the ith class, $i = -1, 1$; and let $f_i(\mathbf{x}) = p_{\mathbf{X}|Y}(\mathbf{x}|y = i)$ be the class-conditional densities. Using the Bayes' rule and the training data, the update on the probabilities of the classes as:

$$q_i(\mathbf{x}) = \Pr(y = i|\mathbf{X} = \mathbf{x}) = \frac{p_{\mathbf{X}|Y}(\mathbf{x}|y = i)\Pr(C_i)}{\sum_j p_{\mathbf{X}|Y}(\mathbf{x}|y = j)\Pr(C_j)} = \frac{f_i(\mathbf{x})p_i}{\sum_j f_j(\mathbf{x})p_j}, \ i = -1, 1. \tag{0.1}$$

The two-class Bayes classifier that equally penalises misclassification can now be defined as:

$$h_B(\mathbf{x}) = \begin{cases} -1, & q_{-1}(\mathbf{x}) > q_1(\mathbf{x}), \\ 1, & \text{otherwise.} \end{cases} \tag{0.2}$$

The decision $q_0(\mathbf{x}) > q_1(\mathbf{x}) \implies p_0 f_0(\mathbf{x}) > p_1 f_1(\mathbf{x})$, and each of the prior, class-conditional pairs describes the joint distribution of $(\mathbf{X}, Y)$. Hence, this is a generative model. The set of points $\{\mathbf{x}|q_0(\mathbf{x}) = q_1(\mathbf{x})\}$ is called the decision boundary. The problem of implementing this classifier is to learn the class-conditional densities and the prior densities from the training data.

1   Bayes' Classifier in $\mathbb{R}^2$

> **Problem (1.1) Bayes' Classifier with Gaussian Class Conditionals**

Let the class conditional distributions be modelled to be multivariate Gaussians. In the training phase, the unknown means and covariances of the Gaussians are estimated from the training data. Let the class conditional densities be parametrised as:

$$f_i(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{(d/2)}(\det \boldsymbol{\Sigma}_i)^{1/2}} e^{(\mathbf{x}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_i)}, \ i = -1, 1. \tag{1.1}$$

The resulting Bayes classifier, using (0.2) results in a quadratic decision boundary as:

$$\frac{1}{2}\mathbf{x}^\mathsf{T}\left(\boldsymbol{\Sigma}_1^{-1}-\boldsymbol{\Sigma}_{-1}^{-1}\right)\mathbf{x}-\left(\mathbf{W}_1-\mathbf{W}_{-1}\right)^\mathsf{T}\mathbf{x}+w_1-w_{-1}=0, \tag{1.2}$$

where $\mathbf{W}_i=\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$ and $w_i=\frac{1}{2}\boldsymbol{\mu}_i^\mathsf{T}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i-\ln\left(p_i\right)+\frac{1}{2}\ln\left(\det\boldsymbol{\Sigma}_i\right)$. The quantities $\frac{1}{2}\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}_i^{-1}\mathbf{x}-\mathbf{W}_i^\mathsf{T}\mathbf{x}+w_i$ are defined to be the score functions for class i, such that the multiclass extension of the 2-class model picks the class with the least score.

To implement this classifier, the means $\boldsymbol{\mu}_i$ and the covariances $\boldsymbol{\Sigma}_i$ are to be estimated from the training data. The maximum likelihood estimators (MLE) for the mean and the covariance, given training samples $\{\mathbf{x}^{(i)},y^{(i)}\}_{i=1}^n$ are given by the sample mean and the sample covariance.

> **Problem (1.2) Bayes' Classifier Trained using Gaussian Mixture Model**

> **Problem (1.3) Bayes' Classifier with Exponential Class Conditional**

2 Bayes' Classifier in $\mathbb{R}^{20}$

> **Problem (2) Bayes' Classifier with Gaussian Class Conditionals**

3 Gaussian Mixture Model in $\mathbb{R}$

> **Problem (3) Bayes' Classifier Trained using Gaussian Mixture Model**

4 Naive Bayes' Classifier for Document Classification