\usepackage[utf8]{inputenc} \usepackage{amsmath} \usepackage{mathrsfs}

Assignment 4

Author

Torin White - 657467127

Q1

[Ex 5 of Chapter 5 of Alpaydin] In addition to Table 5.1, another possibility using Gaussian densities is to have the covariance of $p(x|Ci)$ all diagonal but allow them to be different for different i. Denote the covariance matrix of $p(x|Ci)$ as $diag(s_{i1}^2, s_{i2}^2, \dots, s_{id}^2)$ where diag turns a vector into a diagonal matrix.

The covariance matrix of $p(x|Ci)$ as $diag(s_{i1}^2, s_{i2}^2, \dots, s_{id}^2)$ may be denoted as:

$$\Sigma_i = \begin{bmatrix} S_{i1}^2 & 0 & \dots & 0 \\ 0 & S_{i2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & S_{id}^2 \end{bmatrix}$$

a)

Derive the discriminant gi for this case. (50 points)

With prior probability $P(x|C_1)$ with Guassian distribution $\sim \mathcal{N}_d(\mu_1, \Sigma_1)$ and $P(x|C_2)$ with Guassian distribution $\sim \mathcal{N}_d(\mu_2, \Sigma_2)$

the discriminant function is:

$g_i(x) = P(C_i|X) = P(C_i)P(X|C_i) * C$

taking the log of $g_i(x)$

$\log g_i(x) \triangleq \log P(C_i|X) = \log P(C_i) + \log P(X|C_i) + \log C$

plugging in the density function:

$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$

we get:

$\log g_i(x) \triangleq \log P(C_i) + \log(\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma_i}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_i})^T \boldsymbol{\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})\right))$

For this case of the diagonal covariance matrices:

$$\Sigma_1 = \begin{bmatrix} S_{i1}^2 & 0 & \dots & 0 \\ 0 & S_{i2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & S_{id}^2 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} S_{i1}^2 & 0 & \dots & 0 \\ 0 & S_{i2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & S_{id}^2 \end{bmatrix}$$

find the determinant of Σ_i

$|\Sigma_i| = \prod_{i=1}^{d}(S_{ij}^2)$

$\log g_i(x) \triangleq \log P(C_i) + \log(\frac{1}{(2\pi)^{d/2}|\prod_{i=1}^{d} S_i^2|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})\right))$

find the inverse of Σ_i

$$\Sigma_1^{-1} = \begin{bmatrix} \frac{1}{(S_{i1})^2} & 0 & \dots & 0 \\ 0 & \frac{1}{(S_{i2})^2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{(S_{id})^2} \end{bmatrix}$$

then,

$(\boldsymbol{x}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})$

simplifies to $\sum_{j=1}^{d} \frac{(x_j-\mu_{ij})^2}{(S_{ij})^2}$

then plugging into discriminant function:

$\log g_i(x) \triangleq \log P(C_i) - \frac{d}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{d}\log((S_{ij})^2) - \frac{1}{2}\sum_{j=1}^{d}\frac{(x_j-\mu_{ij})^2}{(S_{ij})^2}$

b)

When does the separating boundary become linear (instead of quadratic)? (50 points)

For the case of linear boundary, set $P(C_1|X) = P(C_2|X)$ or $g_1(x) = g_2(x)$. For the boundary to be linear, the quadratic portion of the equation must cancel. So looking at

$(\boldsymbol{x}-\boldsymbol{\mu_1})^T\boldsymbol{\Sigma_1}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_1}) = (\boldsymbol{x}-\boldsymbol{\mu_2})^T\boldsymbol{\Sigma_2}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_2})$

or

$\sum_{j=1}^{d}\frac{(x_j-\mu_{1j})^2}{S_{1j}^2} = \sum_{j=1}^{d}\frac{(x_j-\mu_{2j})^2}{S_{2j}^2}$

the boundary is linear if $S_{1j}^2 = S_{2j}^2$. Then we are left with:

$\sum_{j=1}^{d}(x_j - \mu_{1j})^2 = \sum_{j=1}^{d}(x_j - \mu_{2j})^2 =$

$\sum_{j=1}^{d} x_j^2 - 2x_j\mu_{1j} + \mu_{1j}^2 = \sum_{j=1}^{d} x_j^2 - 2x_j\mu_{2j} + \mu_{2j}^2 =$

$\sum_{j=1}^{d} -2x_j\mu_{1j} + \mu_{1j}^2 = \sum_{j=1}^{d} -2x_j\mu_{2j} + \mu_{2j}^2$

which leaves no quadratic terms, making the boundary linear.

Q2

[Exercise 5.4 of Alpaydin] But instead of four cases, do it only for the case of $\Sigma 1 \neq \Sigma 2$. You need to derive the expression of $log \frac{P(C_1|x)}{P(C_2|x)}$ using Σ and μ_i, and simplify it as much as possible. There is no need to derive the condition for the boundary to be linear. (60 points)

$\log g_i(x) \triangleq \log P(C_i) + \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma_i}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})\right)$

First, simplify:

$$(\boldsymbol{x}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})$$

=

$$x^T\Sigma_i^{-1}x - \mu_i^T\Sigma_i^{-1}x - x^T\Sigma_i^{-1}\mu_i + \mu_i^T\Sigma_i^{-1}\mu_i$$

Taking the transpose of $\mu_i^T\Sigma_i^{-1}x$ we get $x^T\Sigma_i^{-1}\mu_i$:

$$x^T\Sigma_i^{-1}x - x^T\Sigma_i^{-1}\mu_i - x^T\Sigma_i^{-1}\mu_i + \mu_i^T\Sigma_i^{-1}\mu_i$$

=

$$x^T\Sigma_i^{-1}x - 2x^T\Sigma_i^{-1}\mu_i + \mu_i^T\Sigma_i^{-1}\mu_i$$

the full expression is:

$\log g_i(x) \triangleq \log P(C_i) + \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma_i}|^{1/2}} \exp\left(-\frac{1}{2}(x^T\Sigma_i^{-1}x - 2x^T\Sigma_i^{-1}\mu_i + \mu_i^T\Sigma_i^{-1}\mu_i)\right)$

Now, derive $log \frac{P(C_1|x)}{P(C_2|x)} = \frac{\log g_1(x)}{\log g_2(x)} =$

$$\frac{\log P(C_1)}{\log P(C_2)} + \frac{\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma_1}|^{1/2}} \exp\left(-\frac{1}{2}(x^T\Sigma_1^{-1}x - 2x^T\Sigma_1^{-1}\mu_1 + \mu_1^T\Sigma_1^{-1}\mu_1)\right)}{\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma_2}|^{1/2}} \exp\left(-\frac{1}{2}(x^T\Sigma_2^{-1}x - 2x^T\Sigma_2^{-1}\mu_2 + \mu_2^T\Sigma_2^{-1}\mu_2)\right)}$$

$(2\pi)^{d/2}$ cancels out leaving:

$$\frac{\log P(C_1)}{\log P(C_2)} + \frac{|\boldsymbol{\Sigma_2}|^{1/2} \exp\left(-\frac{1}{2}(x^T\Sigma_1^{-1}x - 2x^T\Sigma_1^{-1}\mu_1 + \mu_1^T\Sigma_1^{-1}\mu_1)\right)}{|\boldsymbol{\Sigma_1}|^{1/2} \exp\left(-\frac{1}{2}(x^T\Sigma_2^{-1}x - 2x^T\Sigma_2^{-1}\mu_2 + \mu_2^T\Sigma_2^{-1}\mu_2)\right)}$$