4.3 NORMALLY DISTRIBUTED PATTERNS

The multivariate normal density function has received considerable attention due to:

(a) Its capability to portray a suitable model for many applications. (b) Being mathematically tractable.

4.3.1 The Univariate Normal Distribution

The scalar normal distribution function given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{(x-\mu)^2}{2\sigma^2}], \quad -\infty < x < \infty \tag{4.3.1}$$

is characterized by two parameters, its *mean*

$$\mu = E[x] = \int_{-\infty}^{\infty} xp(x)dx \tag{4.3.2}$$

and *variance*

$$\sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx \tag{4.3.3}$$

and is frequently denoted by $N(\mu,\sigma^2)$. Simple calculation shows that normally distributed patterns cluster about the mean μ in a way that approximately 68.3% of them fall within the interval $[\mu-\sigma,\mu+\sigma]$, 95.5% within $[\mu-2\sigma,\mu+2\sigma]$ and 99.75 within $[\mu-3\sigma,\mu+3\sigma]$.

4.3.2 The Multivariate Normal Distribution

A generalization of the univariate normal distribution in R^n is given by the multivariate normal distribution function, defined as

$$p(x) = \frac{1}{(2\pi)^{n/2} |C|^{1/2}} \exp[-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)], x \in R^n \tag{4.3.4}$$

where μ is a given vector in R^n and C - an $n \times n$ symmetric positive definite matrix, with inverse C^{-1} and determinant $|C|$. Under these conditions it can be shown that $p(x)$ is a multivariate probability distribution function with mean μ and covariance matrix C, i.e.

$$\mu = E[x] = \int_{R^n} x p(x) dx \tag{4.3.5}$$

$$C = E[(x-\mu)(x-\mu)^T] = \int_{R^n} (x-\mu)(x-\mu)^T p(x) dx \tag{4.3.6}$$

where $dx = dx_1 dx_2 \ldots dx_n$. The elements of μ are

$$\mu_i = \int_{R^n} x_i p(x) dx \tag{4.3.7}$$

while those of C are

$$\sigma_{ij} = \int_{R^n} (x_i - \mu_i)(x_j - \mu_j) p(x) dx \tag{4.3.8}$$

■ **Example 4.3.1** Consider the bivariate normal distribution with its parameters

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad C = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

In this case

$$C^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}, \ |C| = 2$$

and the distribution function is

$$p(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{2}} \exp\left(-\frac{x_1^2}{4} - \frac{x_2^2}{2}\right)$$

The first element of μ is

$$\mu_1 = \frac{1}{2\pi\sqrt{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 \exp\left(-\frac{x_1^2}{4} - \frac{x_2^2}{2}\right) dx_1 dx_2$$

and since x_1 is an odd function we get $\mu_1 = 0$ and similarly $\mu_2 = 0$, i.e. $\mu = 0$. Once μ is known the covariance matrix of $p(\boldsymbol{x})$ can be calculated. In particular

$$\sigma_{11} = \frac{1}{2\pi\sqrt{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^2 \exp\left(-\frac{x_1^2}{4} - \frac{x_2^2}{2}\right) dx_1 dx_2$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x_2^2}{2}\right) dx_2 \cdot \frac{1}{\sqrt{4\pi}} \int_{-\infty}^{\infty} x_1^2 \exp\left(-\frac{x_1^2}{4}\right) dx_1$$

The first integral clearly equals to 1. By using integration by parts, the second integral is replaced by

$$-\frac{1}{\sqrt{4\pi}} 2x_1 \exp\left(-\frac{x_1^2}{4}\right) \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{4\pi}} \int_{-\infty}^{\infty} 2\exp\left(-\frac{x_1^2}{4}\right) dx_1$$

The first part obviously vanishes and by substituting $t = \frac{x_1}{\sqrt{2}}$, the second part becomes

$$\frac{\sqrt{2}}{\sqrt{4\pi}}\int_{-\infty}^{\infty} 2\exp\left(-\frac{t^2}{2}\right)\mathrm{dt} = 2\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right)dt\right] = 2$$

Therefore, $\sigma_{11} = 2$ as expected.

♠

The multivariate normal distribution is determined by $n+\frac{1}{2}n(n+1)$ parameters which are μ_i, $1 \le i \le n$ and σ_{ij}, $1 \le i, j \le n$; $i \le j$. Patterns which are known to be normally distributed create a cluster with center at $\boldsymbol{\mu}$. The shape of this cluster is determined by the covariance matrix $\boldsymbol{C}$. Since $\boldsymbol{C}$ is symmetric positive definite matrix, so is $\boldsymbol{C}^{-1}$ and the equation

$$(\boldsymbol{x}-\boldsymbol{\mu})^T C^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \text{const}$$

is a hyperellipsoid. Thus, the points in R^n with constant probability density are hyperellipsoids whose principal axes are determined by the eigenvectors of C and their lengths - by its eigenvalues.

4.3.3 A Multiclass Multivariate Normal Distribution Problem

Consider a multiclass pattern recognition problem with pattern classes $C_1, C_2, \ldots, C_m$ in R^n, associated with conditional probability distributions

$$p(x \mid C_i) = \frac{1}{(2\pi)^{n/2} \mid \boldsymbol{C}_i \mid^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{C}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right], \boldsymbol{x} \in R^n \quad (4.3.9)$$

for all $1 \le i \le m$. We assume an identity loss matrix and consequently get the decision functions of Eq. (4.2.11) implemented as in Eq. (2.2.6). For all practical purposes one can use $\ln[d_i(\boldsymbol{x})]$, $1 \le i \le m$ instead of $d_i(\boldsymbol{x})$, $1 \le i \le m$. Indeed, $\ln(t)$ is a monotonic increasing function of *t*, i.e.

$$d_i(\boldsymbol{x}) > d_j(\boldsymbol{x}) \text{ if and only if } \ln[d_i(\boldsymbol{x}) > \ln[d_j(\boldsymbol{x})] \quad (4.3.10)$$

Since $p(\boldsymbol{x}|C_i)$ and therefore $d_i(\boldsymbol{x})$ includes an exponential function in its expression, it is convenient to redefine the decision functions as

$$d_i(\boldsymbol{x}) = \ln\left[p(\boldsymbol{x}|C_i)p(C_i)\right] = \ln\left[p(\boldsymbol{x}|C_i)\right] + \ln(p(C_i)) \ , \ 1 \le i \le m \qquad (4.3.11)$$

By substituting the right-hand side of Eq. (4.3.9) in Eq. (4.3.11) we get

$$d_i(\boldsymbol{x}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{C}_i| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\boldsymbol{C}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) + \ln(p(C_i))$$

and since the $i-$independent constant $(-\frac{n}{2}\ln(2\pi))$ can be removed, the decision functions may be taken as

$$d_i(\boldsymbol{x}) = -\frac{1}{2}\ln|\boldsymbol{C}_i| + \ln(p(C_i)) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\boldsymbol{C}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) \qquad (4.3.12)$$

Thus, if the loss matrix is the identity matrix and the patterns are normally distributed, no decision functions will produce better results than the quadratic surfaces given by Eq. (4.3.12).

Quite frequently all the covariance matrices $\boldsymbol{C}_i$ are equal, i.e.

$$\boldsymbol{C}_i = \boldsymbol{C} \ , \ 1 \le i \le m$$

and by removing the new $i-$independent terms one can simplify the decision functions and get

$$d_i(\boldsymbol{x}) = \ln(p(C_i)) + \boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_i \ , 1 \le i \le m \qquad (4.3.13)$$

i.e., linear decision functions (hyperplanes). If we further assume that all the components of $\boldsymbol{x}$ are independent, i.e. $\sigma_{jk} = 0$, $j \ne k$ and that $\sigma_j^2 = 1$, $1 \le j \le n$ then $\boldsymbol{C}$ is the identity matrix of order n and if also $p(C_i) = 1/m$, $1 \le i \le m$ we can remove the constant $\ln(1/m)$ from Eq. (4.3.13) and get

$$d_i(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\mu}_i - \frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i \ , 1 \le i \le m \qquad (4.3.14)$$

which is identical to Eq. (3.2.4) that was derived for classification using the minimum-distance classifier in the case of single prototypes.

The decision boundaries obtained from Eq. (4.3.13) are

$$d_{ij}(\boldsymbol{x}) = d_i(\boldsymbol{x}) - d_j(\boldsymbol{x}) = \ln(p(C_i)) - \ln(p(C_j)) + \boldsymbol{x}^T \boldsymbol{C}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) - \frac{1}{2}\boldsymbol{\mu}_i^T \boldsymbol{C}^{-1}\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_j^T \boldsymbol{C}^{-1}\boldsymbol{\mu}_j, 1 \le i, j \le m \tag{4.3.15}$$

i.e. hyperplanes. If the covariance matrices $\boldsymbol{C}_i$ are not the same, the decision boundaries are quadratic surfaces.

■ **Example 4.3.2** Consider a 2-D 3-class normal distribution problem with covariance matrices

$$\boldsymbol{C}_1 = \boldsymbol{C}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad \boldsymbol{C}_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

mean vectors

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{\mu}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{\mu}_3 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

and $p(C_1) = p(C_2) = 1/4$, $p(C_3) = 1/2$. Thus,

$$|\boldsymbol{C}_1| = |\boldsymbol{C}_2| = 2, \quad \frac{1}{2}\ln|\boldsymbol{C}_1| = \frac{1}{2}\ln|\boldsymbol{C}_2| = \frac{1}{2}\ln 2; \quad |\boldsymbol{C}_3| = 1, \quad \frac{1}{2}\ln|\boldsymbol{C}_3| = 0$$

$$\boldsymbol{C}_1^{-1} = \boldsymbol{C}_2^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad \boldsymbol{C}_3^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the decision functions obtained by Eq. (4.3.12) are

$$d_1(\boldsymbol{x}) = -\frac{1}{2}\ln 2 - \ln 4 - \frac{1}{2}(x_1 - 1, x_2)\begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix}$$

$$= -2.5\ln 2 - \frac{1}{2}\left[(x_1 - 1)^2 + \frac{1}{2}x_2^2\right]$$

$$d_2(\boldsymbol{x}) = -2.5\ln 2 - \frac{1}{2}\left[x_i^2 + \frac{1}{2}(x_2 - 1)^2\right]$$

$$d_3(\boldsymbol{x}) = -\ln 2 - \frac{1}{2}(x_1 - 2, x_2 - 2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 - 2 \\ x_2 - 2 \end{pmatrix}$$

$$= -\ln 2 - \frac{1}{2}\left[(x_1 - 2)^2 + (x_2 - 2)^2\right]$$

The decision boundaries are the straight line

$$d_{12}(\boldsymbol{x}) = d_1(\boldsymbol{x}) - d_2(\boldsymbol{x}) = x_1 - \frac{x_2}{2} - \frac{1}{4} = 0$$

between C_1 and C_2, and the parabolas

$$d_{13}(\boldsymbol{x}) = d_1(\boldsymbol{x}) - d_3(\boldsymbol{x}) = \frac{x_2^2}{4} - 2x_2 - x_1 + \frac{7 - 3\ln 2}{2} = 0$$

$$d_{23}(\boldsymbol{x}) = d_2(\boldsymbol{x}) - d_3(\boldsymbol{x}) = \frac{x_1^2}{4} - 2x_1 - x_2 + \frac{7 - 3\ln 2}{2} = 0$$

between C_1, C_3 and between C_2, C_3 respectively.

♠

4.3.4 Error Probabilities

We will now discuss the error probability associated with the Bayes classifier for normally distributed patterns.

Consider a 2-class pattern recognition problem where the patterns of both classes share the same covariance matrix. Let the multivariate normal densities be

$$p(\boldsymbol{x} \mid C_i) = \frac{1}{(2\pi)^{n/2} \mid \boldsymbol{C} \mid^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \mu_i)^T \boldsymbol{C}^{-1}(\boldsymbol{x} - \mu_i) \right], 1 \le i \le 2 \quad (4.3.16)$$

As previously stated we can simplify the discussion and replace the likelihood ratio $s_{12}(\boldsymbol{x})$ by

$$t_{12}(\boldsymbol{x}) = \ln[s_{12}(\boldsymbol{x})] = \ln[p(\boldsymbol{x} \mid C_1)] - \ln[p(\boldsymbol{x} \mid C_2)] \quad (4.3.17)$$

By virtue of Eq. (4.3.16) we obtain

$$t_{12}(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \mu_2)^T \boldsymbol{C}^{-1}(\boldsymbol{x} - \mu_2) - \frac{1}{2}(\boldsymbol{x} - \mu_1)^T \boldsymbol{C}^{-1}(\boldsymbol{x} - \mu_1) \quad (4.3.18)$$

and since $\boldsymbol{C}^{-1}$ is symmetric this leads to

$$t_{12}(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{C}^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)^T \boldsymbol{C}^{-1}(\mu_1 - \mu_2) \quad (4.3.19)$$

A commonly used 2x2 loss matrix is

$$\begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (4.3.20)$$

for which the threshold value of Eq. (4.2.8) is

$$\lambda = \frac{p(C_2)}{p(C_1)} \quad (4.3.21)$$

Thus, in order to get minimum probability for misclassification one should classify $\boldsymbol{x} \in C_1$ if and only if

$$t_{12}(\boldsymbol{x}) > \ln\left[\frac{p(C_2)}{p(C_1)}\right] \tag{4.3.22}$$

and classify $\boldsymbol{x} \in C_2$ otherwise. Since $\boldsymbol{x}$ is normally distributed and since $t_{12}(\boldsymbol{x})$ is a linear combination of the components of $\boldsymbol{x}$, it must also be normally distributed. The expected value of $t_{12}(\boldsymbol{x})$ with respect to C_1 is (Eq. (4.3.9))

$$\begin{aligned} E_1(t_{12}(\boldsymbol{x})) &= \mu_1 \boldsymbol{C}^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)^T \boldsymbol{C}^{-1}(\mu_1 - \mu_2) \\ &= \frac{1}{2}(\mu_1 - \mu_2)^T \boldsymbol{C}^{-1}(\mu_1 - \mu_2) = \bar{t}_{12}(\boldsymbol{x}) \end{aligned} \tag{4.3.23}$$

The scalar

$$D_{12} = (\mu_1 - \mu_2)^T \boldsymbol{C}^{-1}(\mu_1 - \mu_2) \tag{4.3.24}$$

is called the *Mahalanobis distance* between the distributions $p(\boldsymbol{x}|C_1)$ and $p(\boldsymbol{x}|C_2)$

By definition, the variance of $t_{12}(\boldsymbol{x})$ with respect to C_1 is

$$V_1(t_{12}) = E_1[(t_{12} - \bar{t}_{12}^{\,2})] \tag{4.3.25}$$

From Eqs. (4.3.19) and (4.3.23) we get

$$\begin{aligned} t_{12} - \bar{t}_{12} &= [\boldsymbol{x}^T - \frac{1}{2}(\mu_1 + \mu_2)^T - \frac{1}{2}(\mu_1 - \mu_2)^T]\boldsymbol{C}^{-1}(\mu_1 - \mu_2) \\ &= (\boldsymbol{x} - \mu_1)^T \boldsymbol{C}^{-1}(\mu_1 - \mu_2) \end{aligned} \tag{4.3.26}$$

which implies

$$E_1[(t_{12} - \bar{t}_{12})^2] = E_1\left[(\mu_1 - \mu_2)^T \boldsymbol{C}^{-1}(\boldsymbol{x} - \mu_1)(\boldsymbol{x}^T - \mu_1^T)\boldsymbol{C}^{-1}(\mu_1 - \mu_2)\right] \tag{4.3.27}$$

By virtue of Eq. (4.3.6) we therefore have

$$E_1[(t_{12}-\bar{t}_{12})^2]=(\mu_1-\mu_2)^T C^{-1} C C^{-1}(\mu_1-\mu_2)$$
$$=(\mu_1-\mu_2)^T C^{-1}(\mu_1-\mu_2)=D_{12} \qquad (4.3.28)$$

Thus, $t_{12}(\boldsymbol{x})$, $\boldsymbol{x}\in C_1$ is distributed normally with mean $\frac{D_{12}}{2}$ and variance D_{12}. Similarly, $t_{12}(\boldsymbol{x})$, $\boldsymbol{x}\in C_2$ has a normal distribution with mean $-\frac{D_{12}}{2}$ and variance D_{12}. Consequently

$$p(t_{12}>\alpha \mid C_2)=\frac{1}{\sqrt{2\pi D_{12}}}\int_{\alpha}^{\infty} \exp\left[-\frac{(t_{12}+D_{12}/2)^2}{2D_{12}}\right]dt_{12}$$
$$=1-erf\left(\frac{\alpha+D_{12}/2}{\sqrt{D_{12}}}\right) \qquad (4.3.29)$$

$$p(t_{12}<\alpha \mid C_1)=\frac{1}{\sqrt{2\pi D_{12}}}\int_{-\infty}^{\alpha} \exp\left[-\frac{(t_{12}-D_{12}/2)^2}{2D_{12}}\right]dt_{12}$$
$$=erf\left(\frac{\alpha-D_{12}/2}{\sqrt{D_{12}}}\right) \qquad (4.3.30)$$

where

$$erf(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp(-t^2/2)\,dt \qquad (4.3.31)$$

The error probability to misclassify an arbitrary $\boldsymbol{x}$ is

$$p(error) = p(C_1)p[(t_{12}(\boldsymbol{x}) < \alpha_0 | C_1)] + p(C_2)p[(t_{12}(\boldsymbol{x}) > \alpha_0 | C_2)]$$

$$= p(C_1)erf\left(\frac{\alpha_0 - D_{12}/2}{\sqrt{D_{12}}}\right) + p(C_2)\left[1 - erf\left(\frac{\alpha_0 + D_{12}/2}{\sqrt{D_{12}}}\right)\right] \quad (4.3.32)$$

where

$$\alpha_0 = \ln\left[\frac{p(C_2)}{p(C_1)}\right] \quad (4.3.33)$$

In the particular case $p(C_1) = p(C_2) = \frac{1}{2}$ we get $\alpha_0 = 0$, i.e.

$$p(error) = \frac{1}{2}\left[erf\left(-\sqrt{D_{12}}/2\right) + 1 - erf\left(\sqrt{D_{12}}/2\right)\right]$$

which yields

$$p(error) = \frac{1}{\sqrt{2\pi}}\int_{\sqrt{D_{12}}/2}^{\infty} \exp\left(-t^2/2\right)dt \quad (4.3.34)$$

The quantity D_{12} is the Mahalanobis distance between the distributions $p(\boldsymbol{x}|C_1)$ and $p(\boldsymbol{x}|C_2)$. When this distance increases the error probability decreases, and converges to zero if $D_{12} \to \infty$.

PROBLEMS

1. Consider a 2-D 2-class classification problem, where the patterns of either class are normally distributed with the same covariance matrix

$$\boldsymbol{C} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

The mean vectors of classes C_1 and C_2 are

$$\mu_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

respectively and $p(C_1) = p(C_2) = 1/2$. Get the decision boundary between the two classes.

2. Find the decision boundaries for the following 2-D 3-class classification problem with normally distributed patterns:

$$\boldsymbol{C}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{C}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{C}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\mu_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mu_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \mu_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$p(\boldsymbol{C}_1) = p(\boldsymbol{C}_2) = p(\boldsymbol{C}_3) = \frac{1}{3}$$

3. Find the error probability of the Bayes classifier applied for Problem 1.
4. In Problem 1 choose $p(C_1) = \alpha$, $p(C_2) = 1 - \alpha$ and draw the error probability as a function of α.
5. Consider a 2-D 2-class classification problem with normally distributed patterns. Assume that the vector patterns $(0,0)^T$, $(1,0)^T$, $(0,1)^T$, $(1,1)^T$ belong to C_1 and $(-1,0)^T$, $(0,-1)^T$, $(-1,-1)^T$, $(-2,-2)^T$ to C_2. Approximate μ_i, $\boldsymbol{C}_i$, $i = 1,2$ using *only* these classified patterns and use the results to obtain the decision boundary between the classes.