2. Mathematical Preliminaries

Mathematical Preliminaries

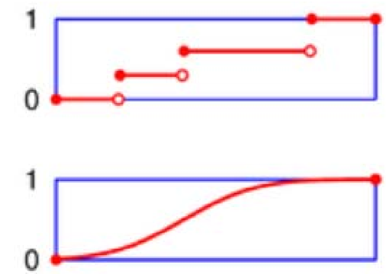
- Random Variables
- Linear Transformations
- Eigenvalues and Eigenvectors
- Orthonormal Transformations
- Matrix Inversion

Random Variables

In statistical pattern recognition, *a pattern* (input to a PR system) is a d-dimensional *feature vector* (*random variable*)

$$\mathbf{x} = (X_1, X_2, ..., X_d)^T$$

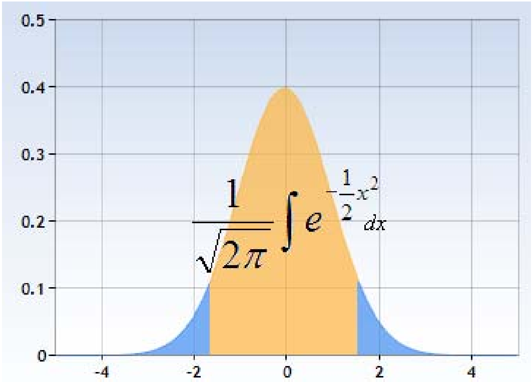
Random Variables



Is fully characterized by its (cumulative) distribution function:

$$F_{\mathbf{X}}(x_1, ..., x_n) = Pr\{X_1 \leq x_1, ..., X_d \leq x_d\}$$

or density function:



$$p_{\mathbf{X}}(x_1, ..., x_d) = \partial^d F(x_1, ..., x_d)/\partial x_1 ... \partial x_d$$

Random Variables

- **In pattern recognition, we deal with random vectors drawn from different classes**
- **Conditional density of class *i* (L classes):**

$$p(\mathbf{x}|\omega_i) = p_i(\mathbf{x}),\ (i = 1, ..., L)$$

- **Unconditional density function of x (mixture density function):**

$$p(\mathbf{x}) = \sum_{i=1}^{L} P_i p_i(\mathbf{x})$$

P_i **is a priori probability of class i**

Random Variables

- A posteriori probability of w_i given x (Bayes theorem):

$$P(\omega_i|\mathbf{x}) = q_i(\mathbf{x}) = \frac{P_i p_i(\mathbf{x})}{p(\mathbf{x})}$$

Random Variables

- **Expected vector (mean):**

$$\mathbf{X} = (X_1, X_2, ..., X_d)^T$$

$$M = E(\mathbf{X}) = (\mu_1, \mu_2, ..., \mu_d)^T$$

Random Variables

- **Covariance matrix (indicates the dispersion of the distribution):**

$$\Sigma = E\{(\mathbf{X}-M)(\mathbf{X}-M)^T\} = \begin{bmatrix} c_{11} & \cdots & c_{1d} \\ & \vdots & \\ c_{d1} & \cdots & c_{dd} \end{bmatrix}$$

$$c_{ii} = \sigma_i^2, \qquad c_{ij} = \rho_{ij}\sigma_i\sigma_j$$

σ_i^2 is the variance, σ_i is the standard deviation of $\mathbf{x}_i$

ρ_{ij} is the correlation coefficient between $\mathbf{x}_i$ and $\mathbf{x}_j$

Random Variables

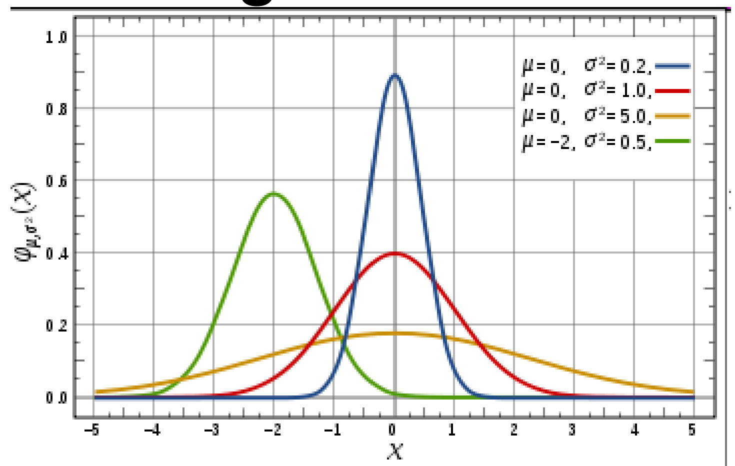
$$\Sigma = E\{\mathbf{X}\mathbf{X}^T\} - E\{\mathbf{X}\}M^T - ME\{\mathbf{X}^T\} + MM^T$$

$$= E\{\mathbf{X}\mathbf{X}^T\} - MM^T = S - MM^T$$

$E\{\mathbf{X}\mathbf{X}^T\} = S$ is the autocorrelation function

Random Variables

Gaussian (normal) distribution: describes data that cluster around a mean or average



1-d

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

d>=1

$$p(\mathbf{x}|M,\Sigma) = \frac{1}{(\sqrt{2\pi})^d \sqrt{|\Sigma|}} \exp\{-\frac{1}{2}\underbrace{(\mathbf{x}-M)^T\Sigma^{-1}(\mathbf{x}-M)}_{\text{distance function}}\}$$

Random Variables

Normal distribution:

- **is uniquely characterized by the expected vector and covariance matrix**

- **The assumption of normality is a reasonable approximation for many real data sets**

- **However, normality should not be assumed without good justification**

- **More about normal distribution**

Linear Transformations

$$\mathbf{y} = A\mathbf{x} \Rightarrow \begin{array}{c} y_1 = a_{11}x_1 + \ldots + a_{1d}x_d \\ \vdots \\ y_d = a_{d1}x_1 + \ldots + a_{dd}x_d \end{array}$$

$$\Leftrightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} = \begin{bmatrix} a_{11} & \ldots & a_{1d} \\ & \vdots & \\ a_{d1} & \ldots & a_{dd} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

Linear Transformations

$$M_{\mathbf{y}} = E\{\mathbf{y}\} = E\{A\mathbf{x}\} = AE\{\mathbf{x}\} = AM_{\mathbf{x}}$$

$$\boxed{\Rightarrow M_{\mathbf{y}} = AM_{\mathbf{x}}}$$

$$\Sigma_{\mathbf{y}} = E\{\underbrace{(\mathbf{y} - M_{\mathbf{y}})}_{A(\mathbf{x} - M_{\mathbf{x}})}\underbrace{(\mathbf{y} - M_{\mathbf{y}})^T}_{(\mathbf{x} - M_{\mathbf{x}})^T A^T}\} = A\Sigma_{\mathbf{x}}A^T$$

Eigenvalues and Eigenvectors

$$Q\phi_i = \lambda_i \phi_i$$

Q: $d \times d$

$\phi_i, i = 1, ..., d$ - eigenvectors

λ_i - eigenvalues

Write for all i:

$$Q[\phi_1, ..., \phi_d] = [\lambda_1\phi_1, ..., \lambda_d\phi_d]$$

$$= [\phi_1, ..., \phi_d] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}$$

Eigenvalues and Eigenvectors

$$Q[\phi_1, ..., \phi_d] = [\lambda_1 \phi_1, ..., \lambda_d \phi_d]$$

$$= [\phi_1, ..., \phi_d] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}$$

eigenvector matrix

$$Q\Phi = \Phi\Lambda$$

eigenvalue matrix

Eigenvalues and Eigenvectors

$$Q\Phi = \Phi\Lambda$$

The eigenvectors corresponding to two different eigenvalues are orthogonal:

$$\phi_i^T \phi_j = 0, \lambda_i \neq \lambda_j$$

$$\Rightarrow \Phi^T \Phi = I$$

$$\Rightarrow \Phi^T = \Phi^{-1}$$

$$\Phi^T Q \Phi = \Lambda$$

Eigenvalues and Eigenvectors

$$\Phi^T \Sigma_{\mathbf{x}} \Phi = \Lambda$$

$$\rho_{ij} = 0$$

$$\mathbf{y} = \Phi^T \mathbf{x}$$

$$\Phi^T \Sigma_{\mathbf{x}} \Phi = \Sigma_{\mathbf{y}} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}$$

Eigenvalues and Eigenvectors

➔ There is no correlation in the transformed space:

$$\mathbf{y} = \Phi^T \mathbf{x} = \begin{bmatrix} \phi_1^T \\ \vdots \\ \phi_d^T \end{bmatrix} \mathbf{x}$$

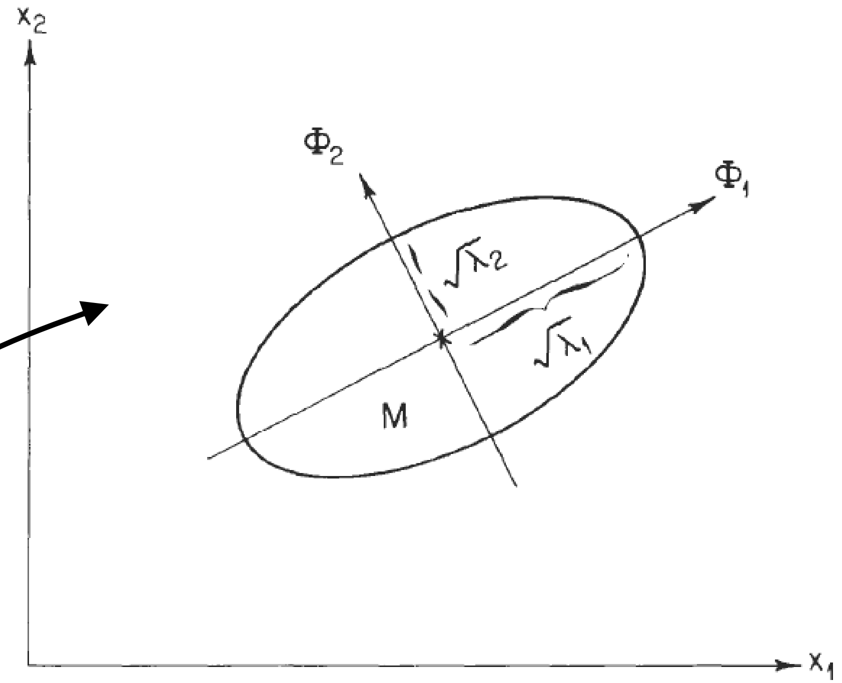
Eigenvalues and Eigenvectors

Eigenvectors are the principal axis of the distribution

$$y_1 = \phi_1^T \mathbf{x}$$

$$\vdots$$

$$y_d = \phi_d^T \mathbf{x}$$



- We rotate the coordinate system
- There is no correlation between y_1 and y_2
- λ_1 gives the variance of y_1, λ_2 gives the variance of y2

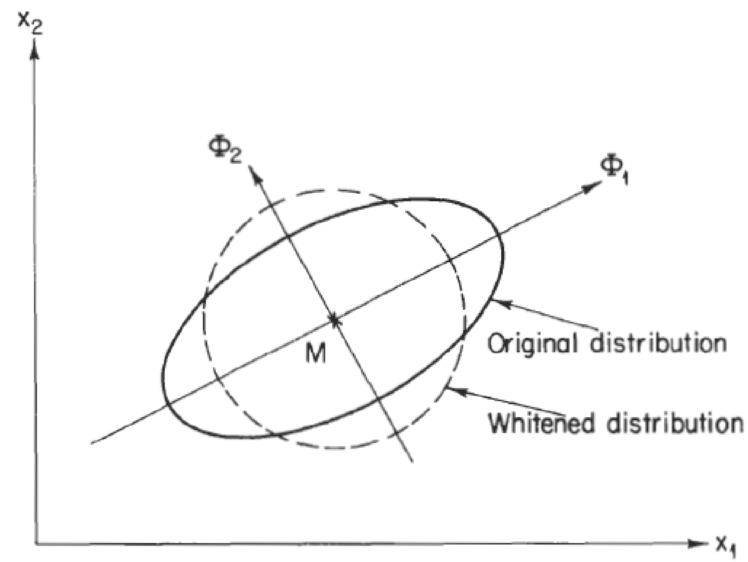
Whitening

After applying the orthonormal transformation, we can add another transformation $\Lambda^{-1/2}$ that will make the covariance matrix equal to *I*:

$$\mathbf{y} = \Lambda^{-1/2}\Phi^T\mathbf{x}$$

$$\Sigma_{\mathbf{y}} = \Lambda^{-1/2}\Phi^T\Sigma_{\mathbf{x}}\Phi\Lambda^{-1/2}$$

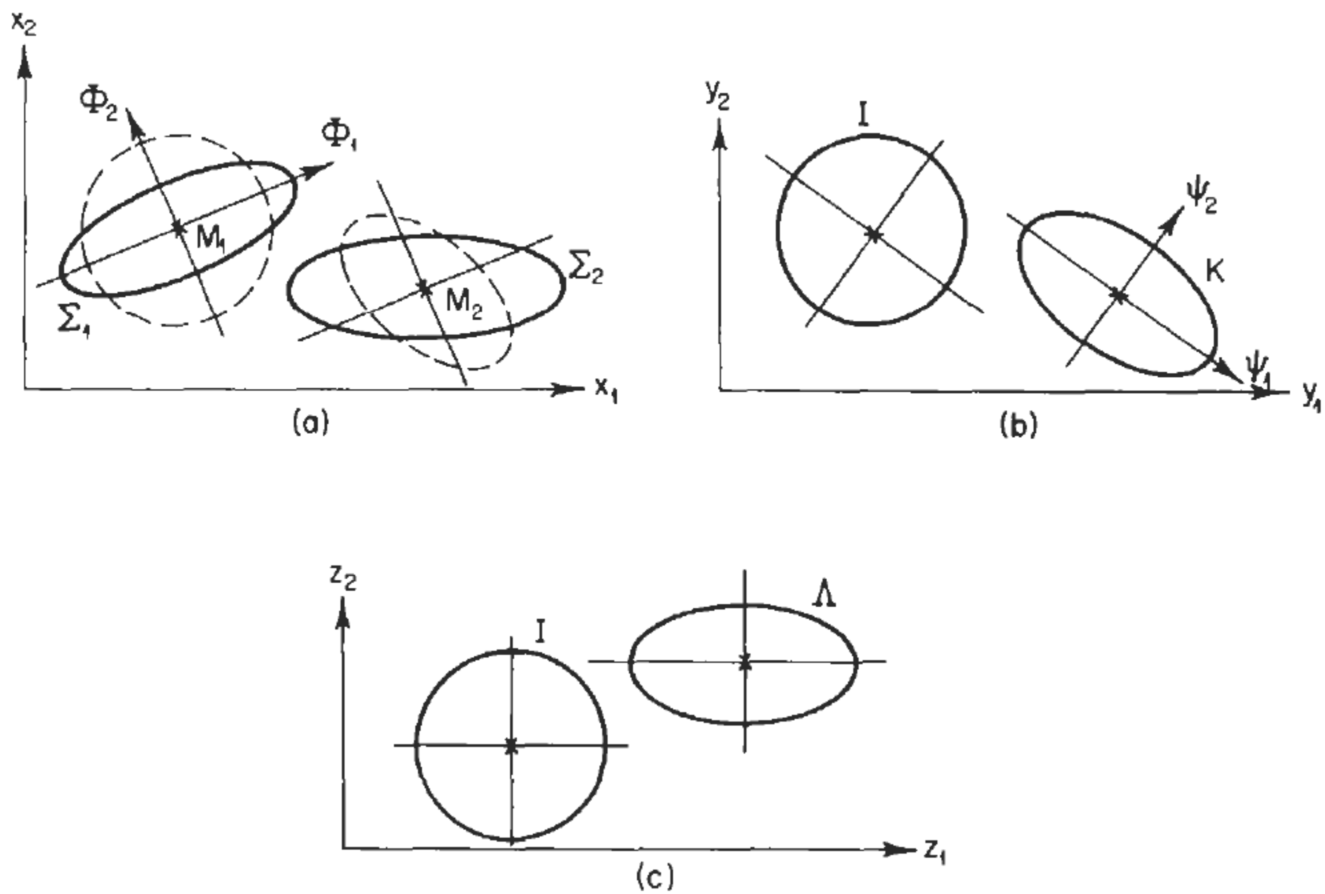
$$= \Lambda^{-1/2}\Lambda\Lambda^{-1/2} = I$$



Purpose: to change the scales of principal components in proportion to $1/\sqrt{\lambda_i}$

Simultaneous Diagonalization

Goal: diagonalize Σ_1 and Σ_2 simultaneously by a linear transformation:



Orthonormal Transformations

$$\mathbf{y} = A\mathbf{x} \qquad AA^T = I$$

Properties:

a) Eigenvalues are invariant

$$\Sigma_{\mathbf{x}}\Phi = \Phi\Lambda$$

$$\Sigma_{\mathbf{y}}\Psi = \Psi K$$

$$\mathbf{y} = A\mathbf{x} \Rightarrow A\Sigma_{\mathbf{x}}A^T = \Sigma_{\mathbf{y}}$$

$$\Rightarrow A\Sigma_{\mathbf{x}}A^T\Psi = \Psi K$$

Orthonormal Transformations

$$\Rightarrow \Sigma_{\mathbf{x}} \underbrace{A^T \Psi}_{\Phi} = \underbrace{A^T \Psi}_{\Phi} \underbrace{K}_{\Lambda}$$

$$\boxed{\Sigma_{\mathbf{x}} \Phi = \Phi \Lambda}$$

Orthnormal Transformations

b) Euclidean distance is invariant

$$\mathbf{x}^T\mathbf{x} = \sum_{i=1}^{d} x_i^2$$

$$\mathbf{y}^T\mathbf{y} = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T\mathbf{x}$$

➔ **Distance in Y-space is the same as in X-space**

Trace of Covariance Matrix

$\mathrm{tr}\ \Sigma_{\mathbf{x}} =$ **Sum of diagonal terms of** $\Sigma_{\mathbf{x}}$

Let's look at $\Phi\Lambda\Phi^T$

$$\mathrm{tr}\ \Phi\Lambda\Phi^T = \mathrm{tr}\ \Lambda\Phi^T\Phi = \mathrm{tr}\ \Lambda$$
$$= \sum_{i=1}^{d} \lambda_i = \mathrm{tr}\ \Sigma_{\mathbf{x}}$$

➔ **The trace of** $\Sigma_{\mathbf{x}}$ **is the summation of all eigenvalues and is invariant under any orthonormal transformation**

Determinant of Covariance Matrix

$$|AB| = |A||B|$$

$$|\Sigma_{\mathbf{x}}| = |\Phi\Lambda\Phi^T| = |\Phi||\Lambda||\Phi^T| =$$
$$|\Lambda||\Phi^T||\Phi| = |\Lambda||\Phi^T\Phi| = |\Lambda||I| = |\Lambda| = \prod_{i=1}^{d} \lambda_i$$

➔ The determinant of $\Sigma_{\mathbf{x}}$ is equal to the product of all eigenvalues and is invariant under any orthonormal transformation

Rank of Covariance Matrix

Rank of $\Sigma_{\mathbf{x}}$ is equal to the number of nonzero eigenvalues

$$\Sigma_{\mathbf{x}} = \Phi\Lambda\Phi^T = \sum_{i=1}^{d} \lambda_i \phi_i \phi_i^T$$

Relation between |S| and $|\Sigma|$:

$$|S| = |\Sigma|(1 = M^T \Sigma^{-1} M)$$

Matrix Inversion

Introduction to matrix inversion: 1 and 2

Generalized Inverse (Pseudo Inverse)

$$QQ^{-1} = I$$

$$Q = \sum_{i=1}^{d} \lambda_i \phi_i \phi_i^T$$

$$Q^{-1} = \sum_{j=1}^{d} \frac{1}{\lambda_j} \phi_j \phi_j^T$$

$$I = QQ^{-1} = \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{\lambda_j} \lambda_i \phi_i \phi_i^T \phi_j \phi_j^T = \sum_{i=1}^{d} \frac{\lambda_i}{\lambda_i} \phi_i \phi_i^T = \sum_{i=1}^{d} \phi_i \phi_i^T$$

Generalized Inverse (Pseudo Inverse)

However, if matrix is singular ➔ some eigenvectors are zero

$$Q = \sum_{i=1}^{r} \lambda_i \phi_i \phi_i^T \qquad r < d,\ \lambda_i = 0 \text{ for } i = r+1, ..., d$$

Generalized (pseudo) inverse:

$$Q^* = \sum_{j=1}^{r} \frac{1}{\lambda_j} \phi_j \phi_j^T$$

$$QQ^* = \sum_{i=1}^{r} \phi_i \phi_i^T \neq I$$

because we skip $i = r+1, ..., d$

But: $$\boxed{QQ^*Q = Q}$$