Exercise 1.4.1 1.4.2.

1.4.9. (spectral decomposition) 1.4.22 (correlation).

2.7.3 2.7.6. (Multivariate Normal)

Exercise 1.4.1. *Consider a matrix* $A \in \mathbb{R}^{3\times 3}$ *and a vector* $\mathbf{x} = (1,2,3) \in \mathbb{R}^3$. *Let*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{pmatrix}.$$

Compute $A\mathbf{x}$ *using both interpretations of matrix-vector multiplication: (i) by taking inner products of rows with* $\mathbf{x}$ *, and (ii) as a linear combination of the columns of* A.

Matrix Multiplication:

AB. if $A \in \mathbb{R}^{n\times m}$

$B \in \mathbb{R}^{m\times k}$

$\Rightarrow AB \in \mathbb{R}^{n\times k}$

Au. if $A \in \mathbb{R}^{n\times m}$

$u \in \mathbb{R}^{m\times 1}$

$\Rightarrow Au \in \mathbb{R}^{n\times 1}$

1.1.1 *Matrix-Vector Multiplication* $A\mathbf{x}$

Let $\mathbf{x} \in \mathbb{R}^m$ be a vector, with $\mathbf{x} = (x_1, \ldots, x_m)$, and let $A \in \mathbb{R}^{n\times m}$ be a matrix. Recall that the matrix-vector product $A\mathbf{x}$ is defined by the rule:

$$(A\mathbf{x})_i = \sum_{j=1}^{m} A_{ij} x_j,$$

which states that the i-th entry of the vector $A\mathbf{x} \in \mathbb{R}^n$ is the inner product of the i-th row of A with the vector $\mathbf{x}$.

An important alternative interpretation of $A\mathbf{x}$ is that it is a linear combination of the columns of A with coefficients given by the entries of $\mathbf{x}$:

$$A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m, \tag{1.1}$$

where $\mathbf{a}_j \in \mathbb{R}^m$ denotes the j-th column of A.

$$Ax = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1+2\times 2+3\times 3 \\ 0+1\times 2+4\times 3 \\ 5\times 1+6\times 2+0 \end{pmatrix} = \begin{pmatrix} 14 \\ 14 \\ 17 \end{pmatrix}$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ a_1 & a_2 & a_3 \\ \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix} & \begin{pmatrix} 2 \\ 1 \\ 6 \end{pmatrix} & \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix} \end{array}$$

$$Ax = (a_1\ a_2\ a_3)\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = a_1 + 2a_2 + 3a_3 = \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix} + 2\begin{pmatrix} 2 \\ 1 \\ 6 \end{pmatrix} + 3\begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 14 \\ 14 \\ 17 \end{pmatrix}$$

Exercise 1.4.2. *Let* $\mathbf{x} = (1,2,3) \in \mathbb{R}^3$ *and* $\mathbf{y} = (4,5,6) \in \mathbb{R}^3$. *Compute the matrix* $\mathbf{x}\mathbf{y}^\top$. *Argue why this matrix has rank one. Compute* $\mathrm{tr}(\mathbf{x}\mathbf{y}^\top)$ *and* $\mathbf{x}^\top\mathbf{y}$. *Is it a coincidence that these two numbers are equal?*

A **rank-one matrix** is formed by the outer product of two vectors. Specifically, if $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, the outer product $\mathbf{x}\mathbf{y}^\top \in \mathbb{R}^{n\times m}$ is defined as:

$$A = \mathbf{x}\mathbf{y}^\top \quad \text{with} \quad A_{ij} = x_i y_j \quad \text{for all } i = 1,\ldots,n,\ j = 1,\ldots,m.$$

The resulting matrix has rank[2] at most one because all rows (or columns) of the matrix are scalar multiples of each other.

Rank: number of linearly ind. rows/cols.

$$\underset{\mathbb{R}^{3\times1}\ \mathbb{R}^{1\times3}}{xy^T} = \begin{pmatrix}1\\2\\3\end{pmatrix}(4\ 5\ 6) = \begin{pmatrix}4&5&6\\8&10&12\\12&15&18\end{pmatrix}$$

$$\begin{array}{ccc} a_1 & a_2 & a_3 \end{array}$$

$a_2 = \frac{5}{4} a_1$

$a_3 = \frac{3}{2} a_1$ $\Rightarrow$ rank 1.

$$tr(xy^T) = 4 + 10 + 18 = 32.$$

$$\underset{\mathbb{R}^{1\times3}\ \mathbb{R}^{3\times1}}{x^T y} = (1\ 2\ 3)\begin{pmatrix}4\\5\\6\end{pmatrix} = 1\times4 + 2\times5 + 3\times6 = 32.$$

Properties of trace.

① if $\underset{\mathbb{R}^{n\times m}}{A}\ \underset{\mathbb{R}^{m\times n}}{B}$ & BA are both valid.

$tr(AB) = tr(BA)$

② $tr(A) = tr(A^T)$

$A \in \mathbb{R}^{n\times n}$

$$\underline{tr(xy^T)} = tr(y^T x)$$

$$= tr(\underline{x^T y}) = \underline{x^T y}.$$

Exercise 1.4.9. *Let* $A \in \mathbb{S}^2$ *be the matrix*

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}.$$

$\leftarrow$ symmetric.

Compute the eigenvalues and eigenvectors of A *by solving the characteristic equation and verify the spectral theorem by expressing* A *as* $A = U\Lambda U^\top$, *where* U *is an orthogonal matrix of eigenvectors and* Λ *is the diagonal matrix of eigenvalues.*

step 1. $\det(\lambda I - A) = \det\begin{pmatrix} \lambda-3 & -2 \\ -2 & \lambda-6 \end{pmatrix} = (\lambda-3)(\lambda-6) - 4 = 0$

$(\lambda-3)(\lambda-6)-4$

$= \lambda^2 - 9\lambda + 14 = (\lambda-2)(\lambda-7) = 0 \Rightarrow \lambda_1 = 7 \quad \lambda_2 = 2.$

step 2. $A v_1 = \lambda_1 v_1$ assume $v_1 = \begin{pmatrix} a \\ b \end{pmatrix}$

$\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = 7\begin{pmatrix} a \\ b \end{pmatrix} \Rightarrow \begin{cases} 3a+2b = 7a. \\ 2a+6b = 7b \end{cases} \Rightarrow b = 2a.$

$a^2+b^2=1.$

$\Rightarrow a = \frac{1}{\sqrt5} \quad b = \frac{2}{\sqrt5} \quad v_1 = \begin{pmatrix} \frac{1}{\sqrt5} \\ \frac{2}{\sqrt5} \end{pmatrix}$

$A v_2 = \lambda_2 v_2$ assume $v_2 = \begin{pmatrix} c \\ d \end{pmatrix}$

$\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix} = 2\begin{pmatrix} c \\ d \end{pmatrix} \Rightarrow \begin{cases} 3c+2d = 2c \\ 2c+6d = 2d \end{cases} \Rightarrow c = -2d.$

$c^2+d^2=1$

$\Rightarrow c = -\frac{2}{\sqrt5} \quad d = \frac{1}{\sqrt5} \quad v_2 = \begin{pmatrix} -\frac{2}{\sqrt5} \\ \frac{1}{\sqrt5} \end{pmatrix}.$ $\quad U = (v_1 \ v_2).$

$\Lambda = \begin{pmatrix} 7 & 0 \\ 0 & 2 \end{pmatrix} \quad U = \begin{pmatrix} \frac{1}{\sqrt5} & -\frac{2}{\sqrt5} \\ \frac{2}{\sqrt5} & \frac{1}{\sqrt5} \end{pmatrix} \quad A = U \Lambda U^T$

Exercise 1.4.22. *Show that if* $Y = aX + b$ *then* $\operatorname{corr}(X,Y) = \pm 1$.

$a \neq 0.$

$$\operatorname{corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\cdot\sqrt{\operatorname{Var}(Y)}}.$$

$$\operatorname{Cov}(X,Y) = \operatorname{Cov}(X, aX+b) = \operatorname{Cov}(X, aX) = a\operatorname{Cov}(X,X) = a\operatorname{Var}(X).$$

$$\operatorname{Var}(Y) = \operatorname{Var}(aX+b) = \operatorname{Var}(aX) = a^2\operatorname{Var}(X).$$

$$\operatorname{Cov}(aX, aX) = a \times a \times \operatorname{Cov}(X,X).$$

$$\operatorname{Corr}(X,Y) = \frac{a\operatorname{Var}(X)}{\sqrt{\operatorname{Var}(X)}\sqrt{a^2\operatorname{Var}(X)}} = \frac{a\,\cancel{\operatorname{Var}(X)}}{\cancel{\sqrt{\operatorname{Var}(X)}}\cdot|a|\cdot\cancel{\sqrt{\operatorname{Var}(X)}}} = \frac{a}{|a|}$$

$$= \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

Exercise 2.7.3. *Consider two random vectors $X \sim N_m(\mu_X, \Sigma_X)$ and $Y \sim N_q(\mu_Y, \Sigma_Y)$. Show that if X and Y are independent, the joint distribution of (X, Y) is multivariate normal with mean (μ_X, μ_Y) and block diagonal covariance matrix.*

Thus, for $X = \mu + \Sigma^{1/2} Z$, we get:

$$\phi_X(\mathbf{t}) = \mathbb{E}e^{i\mathbf{t}^\top(\mu+\Sigma^{1/2}Z)} = e^{i\mathbf{t}^\top\mu}\mathbb{E}e^{i(\Sigma^{1/2}\mathbf{t})^\top Z} = e^{i\mathbf{t}^\top\mu}\phi_Z(\Sigma^{1/2}\mathbf{t}) = e^{i\mathbf{t}^\top\mu - \frac{1}{2}\mathbf{t}^\top\Sigma\mathbf{t}}, \tag{2.3}$$

which gives the formula for the characteristic function of the multivariate normal distribution for general μ and Σ.

Method 1: p.d.f.

$f_X(x)$ $\quad$ $f_Y(y)$

$X \sim N_m(\mu_X, \Sigma_X)$

$$f_{(X,Y)}(x, y) \overset{\text{ind.}}{=} f_X(x) \cdot f_Y(y) \to \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{m+q}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{pmatrix}\right)$$

Method 2. → characteristic function.

$\triangle$ $t = (\underbrace{t_1, t_2, \cdots t_m}_{t_X}, \underbrace{t_{m+1}, \cdots t_{m+q}}_{t_Y})$

$$\phi_{(X,Y)}(t) = E\left(e^{it^T\binom{X}{Y}}\right)$$

$$= E\left(e^{i(t_X^T X + t_Y^T Y)}\right)$$

$E(XY) \overset{\text{ind.}}{=} E(X)E(Y)$

$$= E\left(e^{it_X^T X + it_Y^T Y}\right) = E\left(\underline{e^{it_X^T X}} \cdot \underline{e^{it_Y^T Y}}\right)$$

$$\overset{\text{ind.}}{=} E\left(e^{it_X^T X}\right) \cdot E\left(e^{it_Y^T Y}\right) = \phi_X(t_X) \cdot \phi_Y(t_Y)$$

$$= e^{it_X^T\mu_X - \frac{1}{2}t_X^T\Sigma_X t_X} \cdot e^{it_Y^T\mu_Y - \frac{1}{2}t_Y^T\Sigma_Y t_Y}$$

$$= e^{i(\underline{t_X^T\mu_X + t_Y^T\mu_Y}) - \frac{1}{2}(\underline{t_X^T\Sigma_X t_X + t_Y^T\Sigma_Y t_Y})}$$

$$= e^{it^T\binom{\mu_X}{\mu_Y} - \frac{1}{2}\underline{t^T\begin{pmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{pmatrix} t}}$$

$$\begin{pmatrix} t_X \\ t_Y \end{pmatrix}^T \begin{pmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{pmatrix} \begin{pmatrix} t_X \\ t_Y \end{pmatrix}$$

Exercise 2.7.6. *Let $X \sim N_m(\mu, \Sigma)$. Use the spectral decomposition of Σ to transform X into independent standard normal variables.*

$\Sigma = U\Lambda U^T$

positive-definite.

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix}$$

$\lambda_1, \lambda_2 \cdots \lambda_m > 0.$

By Proposition 2.1.1, if $Z \sim N_m(\mathbf{0}, I_m)$ then $\mu + \Sigma^{1/2}Z \sim N_m(\mu, \Sigma)$. In this section, we generalize this result. More generally, for any matrix $A \in \mathbb{R}^{p\times m}$ and vector $b \in \mathbb{R}^p$, if $X \sim N_m(\mu, \Sigma)$, then the linear transformation $AX + b$ is distributed as:

conclusion.

$$AX + b \sim N_p(A\mu + b, A\Sigma A^\top). \tag{2.4}$$

$X \sim N_m(\mu, \Sigma) \rightarrow$ ind. s. n.

Step 1. $X - \mu \sim N_m(0, \Sigma)$, $0 \in \mathbb{R}^{m\times 1}$

Step 2. need Σ to be diagnal. so that independent.

$$U^T(X-\mu) \sim N_m(0, U^T \Sigma U)$$

$$\Sigma = U \Lambda U^T$$

$UU^T = U^T U = I_m.$

$$U^T(X-\mu) \sim N_m(0, \Lambda)$$

Step 3.

$$\Lambda^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\lambda_m}} \end{pmatrix}.$$

$$\Lambda^{-\frac{1}{2}} U^T(X-\mu) \sim N_m(0, \Lambda^{-\frac{1}{2}} \Lambda \Lambda^{-\frac{1}{2}})$$

$$N_m(0, I_m).$$