Probability Theory

Lecture 5

Per Sidén

Division of Statistics
Dept. of Computer and Information Science
Linköping University

Overview Lecture 5

- Linear algebra recap
- Multivariate normal distribution

Linear algebra recap

- **Eigen-decomposition** of an $n \times n$ symmetric matrix $\mathbf{A}$

$$\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{D}$$

where $\mathbf{D} = Diag(\lambda_1, ..., \lambda_n)$ and $\mathbf{C}$ is an orthogonal matrix.

Linear algebra recap

- **Eigen-decomposition** of an $n \times n$ symmetric matrix $\mathbf{A}$

$$\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{D}$$

  where $\mathbf{D} = Diag(\lambda_1, ..., \lambda_n)$ and $\mathbf{C}$ is an orthogonal matrix.
- **Orthogonal matrix**:
  - $\mathbf{C}'\mathbf{C} = \mathbf{I}$
  - $\mathbf{C}^{-1} = \mathbf{C}'$
  - $\det \mathbf{C} = \pm 1$

Linear algebra recap

- **Eigen-decomposition** of an $n \times n$ symmetric matrix $\mathbf{A}$

$$\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{D}$$

where $\mathbf{D} = Diag(\lambda_1, ..., \lambda_n)$ and $\mathbf{C}$ is an orthogonal matrix.

- **Orthogonal matrix**:
  - $\mathbf{C}'\mathbf{C} = \mathbf{I}$
  - $\mathbf{C}^{-1} = \mathbf{C}'$
  - $\det \mathbf{C} = \pm 1$
- The columns of $\mathbf{C} = (\mathbf{c}_1, ..., \mathbf{c}_n)$ are the eigenvectors, and λ_i is the ith largest eigenvalue.
- $\det \mathbf{A} = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$.

Quadratic forms and positive-definiteness

- **Quadratic form**
$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$$
- $Q(\mathbf{x})$ is **positive-definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- $Q(\mathbf{x})$ is **positive-semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$.

Quadratic forms and positive-definiteness

- **Quadratic form**
$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$$
- $Q(\mathbf{x})$ is **positive-definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- $Q(\mathbf{x})$ is **positive-semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- $Q(\mathbf{x})$ is **positive-definite** iff all eigenvalues of $\mathbf{A}$ are positive.
- $Q(\mathbf{x})$ is **positive-semidefinite** iff all eigenvalues of $\mathbf{A}$ are non-negative.

Matrix square root

- If $D = diag(\lambda_1, ..., \lambda_n)$ is diagonal, then $\tilde{D} = diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})$ is the square root of D:
$$\tilde{D}\tilde{D} = D$$
and we can write $D^{1/2} = \tilde{D}$.

Matrix square root

- If $D = diag(\lambda_1, ..., \lambda_n)$ is diagonal, then $\tilde{D} = diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})$ is the square root of D:
$$\tilde{D}\tilde{D} = D$$
and we can write $D^{1/2} = \tilde{D}$.
- The **square root of a positive definite matrix** A
$$A = CDC'$$
can be defined as
$$A^{1/2} = C\tilde{D}C'$$
where $\tilde{D} == diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})$.

Matrix square root

- If $D = diag(\lambda_1, ..., \lambda_n)$ is diagonal, then $\tilde{D} = diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})$ is the square root of D:
$$\tilde{D}\tilde{D} = D$$
and we can write $D^{1/2} = \tilde{D}$.
- The **square root of a positive definite matrix** A
$$A = CDC'$$
can be defined as
$$A^{1/2} = C\tilde{D}C'$$
where $\tilde{D} == diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})$.
- Check:
$$A^{1/2}A^{1/2} = C\tilde{D}C'C\tilde{D}C' = C\tilde{D}\tilde{D}C' = CDC' = A$$
- We also have
$$\left(A^{-1}\right)^{1/2} = \left(A^{1/2}\right)^{-1}$$
which is denoted by $A^{-1/2}$.

Covariance matrix

- Mean vector

$$\mu = \mathrm{E}\mathbf{X} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$$

Covariance matrix

- Mean vector

$$\mu = \mathrm{E}\mathbf{X} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$$

- Covariance matrix

$$\Lambda = Cov(\mathbf{X}) = E(\mathbf{X} - \mu)(\mathbf{X} - \mu)'$$

Covariance matrix

- Mean vector

$$\mu = \mathrm{E}\mathbf{X} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$$

- Covariance matrix

$$\Lambda = Cov(\mathbf{X}) = E(\mathbf{X} - \mu)(\mathbf{X} - \mu)'$$

Th Every covariance matrix is positive semidefinite.

- $\det \Lambda \geq 0$.

Linear transformations

- Recall that if $Y = aX + b$, where $E(X) = \mu$ and $Var(X) = \sigma^2$ then

$$E(Y) = a\mu + b$$
$$Var(Y) = a^2\sigma^2$$

Linear transformations

- Recall that if $Y = aX + b$, where $E(X) = \mu$ and $Var(X) = \sigma^2$ then

$$
\begin{aligned}
E(Y) &= a\mu + b \\
Var(Y) &= a^2\sigma^2
\end{aligned}
$$

Th **Multivariate linear transformation**

Let $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$, where $\mathbf{X}$ is $n \times 1$ and $\mathbf{B}$ is $m \times n$.
Assume $E\mathbf{X} = \mu$ and $Cov(\mathbf{X}) = \Lambda$. Then,

$$
\begin{aligned}
E(\mathbf{Y}) &= \mathbf{B}\mu + \mathbf{b} \\
Cov(\mathbf{Y}) &= \mathbf{B}\Lambda\mathbf{B}'
\end{aligned}
$$

Linear transformations

- Recall that if $Y = aX + b$, where $E(X) = \mu$ and $Var(X) = \sigma^2$ then

$$E(Y) = a\mu + b$$
$$Var(Y) = a^2\sigma^2$$

Th **Multivariate linear transformation**

Let $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$, where $\mathbf{X}$ is $n \times 1$ and $\mathbf{B}$ is $m \times n$.
Assume $E\mathbf{X} = \mu$ and $Cov(\mathbf{X}) = \Lambda$. Then,

$$E(\mathbf{Y}) = \mathbf{B}\mu + \mathbf{b}$$
$$Cov(\mathbf{Y}) = \mathbf{B}\Lambda\mathbf{B}'$$

Th Let $\mathbf{X} = (X_1, ..., X_n)'$ where $X_1, ..., X_n \overset{iid}{\sim} N(0, 1)$. Then

$$\mathbf{Y} = \mu + \Lambda^{1/2}\mathbf{X} \sim N(\mu, \Lambda)$$

Multivariate normal distribution

- Multivariate normal $\mathbf{X} \sim N(\mu, \Lambda)$, where $\mathbf{X}$ is a $n \times 1$ random vector.

Multivariate normal distribution

- Multivariate normal $\mathbf{X} \sim N(\mu, \Lambda)$, where $\mathbf{X}$ is a $n \times 1$ random vector.
- Three equivalent definitions:
  - $\mathbf{X}$ is (multivariate) normal iff $\mathbf{a}'\mathbf{X}$ is (univariate) normal for all $\mathbf{a}$.

Multivariate normal distribution

- Multivariate normal $\mathbf{X} \sim N(\mu, \Lambda)$, where $\mathbf{X}$ is a $n \times 1$ random vector.
- Three equivalent definitions:
  - $\mathbf{X}$ is (multivariate) normal iff $\mathbf{a}'\mathbf{X}$ is (univariate) normal for all $\mathbf{a}$.
  - $\mathbf{X}$ is multivariate normal iff its characteristic function is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = Ee^{i\mathbf{t}'\mathbf{X}} = \exp\left(i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Lambda\mathbf{t}\right)$$

Multivariate normal distribution

- Multivariate normal $\mathbf{X} \sim N(\mu, \Lambda)$, where $\mathbf{X}$ is a $n \times 1$ random vector.
- Three equivalent definitions:
  - $\mathbf{X}$ is (multivariate) normal iff $\mathbf{a}'\mathbf{X}$ is (univariate) normal for all $\mathbf{a}$.
  - $\mathbf{X}$ is multivariate normal iff its characteristic function is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = Ee^{i\mathbf{t}'\mathbf{X}} = \exp\left(i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Lambda\mathbf{t}\right)$$

  - $\mathbf{X}$ is multivariate normal iff its density function is of the form

$$f_{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{\det \Lambda}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Lambda^{-1}(\mathbf{x}-\mu)\right\}$$

Multivariate normal distribution

- Multivariate normal $\mathbf{X} \sim N(\mu, \Lambda)$, where $\mathbf{X}$ is a $n \times 1$ random vector.
- Three equivalent definitions:
  - $\mathbf{X}$ is (multivariate) normal iff $\mathbf{a}'\mathbf{X}$ is (univariate) normal for all $\mathbf{a}$.
  - $\mathbf{X}$ is multivariate normal iff its characteristic function is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = Ee^{i\mathbf{t}'\mathbf{X}} = \exp\left(i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Lambda\mathbf{t}\right)$$

  - $\mathbf{X}$ is multivariate normal iff its density function is of the form

$$f_{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{\det \Lambda}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Lambda^{-1}(\mathbf{x}-\mu)\right\}$$

- **Bivariate normal** $(n = 2)$

$$\Lambda = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where $-1 \leq \rho \leq 1$ is the correlation coefficient.

Properties of the normal distribution

- Let $\mathbf{X} \sim N(\mu, \Lambda)$.

TH Linear combinations: $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$, where $\mathbf{X}$ is $n \times 1$ and $\mathbf{B}$ is $m \times n$. Then

$$\mathbf{Y} \sim N(\mathbf{B}\mu + \mathbf{b}, \mathbf{B}\Lambda\mathbf{B}')$$

Properties of the normal distribution

- Let $\mathbf{X} \sim N(\mu, \Lambda)$.

Th Linear combinations: $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$, where $\mathbf{X}$ is $n \times 1$ and $\mathbf{B}$ is $m \times n$. Then

$$\mathbf{Y} \sim N(\mathbf{B}\mu + \mathbf{b}, \mathbf{B}\Lambda\mathbf{B}')$$

Cor The components of $\mathbf{X}$ are all normal ($\mathbf{B} = (0, \ldots 1, 0, \ldots, 0)$)

$$Y_i \sim N(\mu_i, \Lambda_{ii})$$

Properties of the normal distribution

- Let $\mathbf{X} \sim N(\mu, \Lambda)$.

TH Linear combinations: $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$, where $\mathbf{X}$ is $n \times 1$ and $\mathbf{B}$ is $m \times n$. Then

$$\mathbf{Y} \sim N(\mathbf{B}\mu + \mathbf{b}, \mathbf{B}\Lambda\mathbf{B}')$$

COR The components of $\mathbf{X}$ are all normal ($\mathbf{B} = (0, \ldots 1, 0, \ldots, 0)$)

$$Y_i \sim N(\mu_i, \Lambda_{ii})$$

COR Let $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ where $\mathbf{X}_1$ is $n_1 \times 1$ and $\mathbf{X}_2$ is $n_2 \times 1$ ($n_1 + n_2 = n$). Then

$$\mathbf{X}_1 \sim N(\mu_1, \Lambda_1)$$

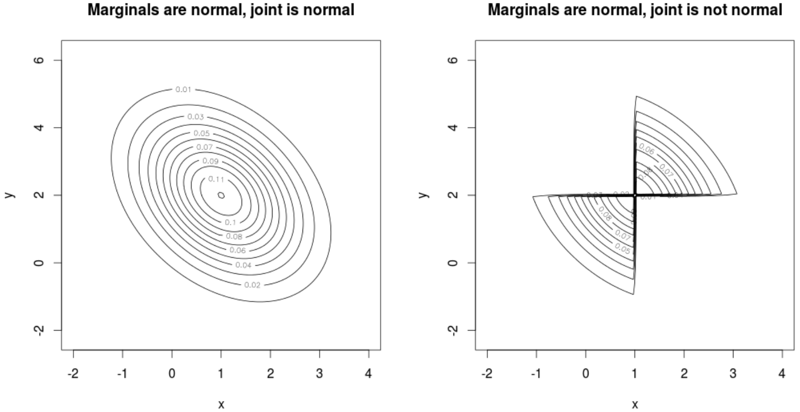
where μ_1 are the n_1 first elements of μ and Λ_1 is the $n_1 \times n_1$ submatrix of Λ.

Marginal normal may not be jointly normal

- We know that $\mathbf{X} \sim N(\mu, \Lambda)$ implies that all marginals are normal.

Marginal normal may not be jointly normal

- We know that $\mathbf{X} \sim N(\mu, \Lambda)$ implies that all marginals are normal.
- The converse does not hold. Normal marginals does not imply that the joint distribution is normal.



Conditional distributions from $N(\mu, \Lambda)$

- Let $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2(\mu, \Lambda)$, where

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

- Then

$$Y|X = x \sim N\left[\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x),\ \sigma_y^2(1 - \rho^2)\right]$$

Conditional distributions from $N(\mu, \Lambda)$

- Let $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2(\mu, \Lambda)$, where

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

- Then

$$Y|X = x \sim N\left[\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x),\ \sigma_y^2(1 - \rho^2)\right]$$

- The **regression function** $E(Y|X)$ is **linear** and $Var(Y|X) =$residual variance.

Conditional distributions from $N(\mu, \Lambda)$

- Let $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2(\mu, \Lambda)$, where

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

- Then

$$Y|X = x \sim N\left[\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x),\ \sigma_y^2(1-\rho^2)\right]$$

- The **regression function** $E(Y|X)$ is **linear** and $Var(Y|X)$ =residual variance.

Th Let $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ and partition μ and Λ accordingly as

$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$. Then

$$\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2 \sim N\left[\mu_1 + \Lambda_{12}\Lambda_{22}^{-1}(x_2 - \mu_2),\ \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right]$$

Independence and normality

- Correlation measures **linear** association (dependence).

Independence and normality

- Correlation measures **linear** association (dependence).
- In general: Uncorrelated $\nrightarrow$ Independence.

Independence and normality

- Correlation measures **linear** association (dependence).
- In general: Uncorrelated $\nrightarrow$ Independence.
- In the normal distribution: Uncorrelated $\leftrightarrow$ Independence.

Independence and normality

- Correlation measures **linear** association (dependence).
- In general: Uncorrelated $\nrightarrow$ Independence.
- In the normal distribution: Uncorrelated $\leftrightarrow$ Independence.
- Remember that: X and Y are jointly normal $\rightarrow$ the regression function is linear $\rightarrow$the linear predictor is optimal.

Independence and normality

- Correlation measures **linear** association (dependence).
- In general: Uncorrelated $\nrightarrow$ Independence.
- In the normal distribution: Uncorrelated $\leftrightarrow$ Independence.
- Remember that: X and Y are jointly normal $\rightarrow$ the regression function is linear $\rightarrow$the linear predictor is optimal.
- $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ are independent.

Principal components

- Let $\mathbf{C}\Lambda\mathbf{C}' = \mathbf{D} = diag(\lambda_1, ..., \lambda_n)$.

Principal components

- Let $\mathbf{C}\Lambda\mathbf{C}' = \mathbf{D} = diag(\lambda_1, ..., \lambda_n)$.

Th Let $\mathbf{X} \sim N(\mu, \Lambda)$ and set $\mathbf{Y} = \mathbf{C}'\mathbf{X}$, then

$$\mathbf{Y} \sim N(\mathbf{C}'\mu, \mathbf{D})$$

so that the components of $\mathbf{Y}$ are independent and $Var(Y_i) = \lambda_i$.