Multivariate Statistics

Sudipta Das

Assistant Professor,
Department of Data Science,
Ramakrishna Mission Vivekananda University, Kolkata
Slides adapted from Jhonson & Winchern

Outline I

1. Review of Linear Algebra
   - Vectors and Matrix
   - Matrix inequalities and Maximization

2. Random Vectors & Random Sample
   - Random Vectors
   - Random Samples
   - Generalized Sample Variance
   - Statistical Distance

Matrix and Random Vectors I

- Vectors

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{bmatrix} \; or \;\; x' = \left[x_1, x_2, x_3, \ldots, x_p\right]$$

- Eucledian distance form origin, length or 2-norm

$$||x||_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_p^2}.$$

- Angle between vectors x and y

$$\cos(\theta) = \frac{x'.y}{||x||_2||y||_2}$$

Matrix and Random Vectors II

- Linear dependence of vectors:- A set of vectors $x_1, x_2, \ldots, x_n$ is said to be linearly dependent if there exists constants $c_1, c_2, \ldots, c_k$, not all zero, such that

$$c_1 x_1 + c_2 x_2 + \ldots + c_n x_n = 0.$$

- Vectors of same dimensions that are not linearly dependent are said to be linearly independent.

Matrix and Random Vectors III

- Matrices

$$A_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

- A square matrix $A_{n\times n}$ is symmetric if $A = A'$.
- Inverse of a square matrix A is A^{-1}, where $|A| \neq 0$ and $AA^{-1} = I$.
- A matrix Q is called orthogonal matrix if

$$Q^{-1} = Q'.$$

- A square matrix A is said to have an eigenvalue λ, with corresponding eigenvector $x \neq 0$, if

$$Ax = \lambda x.$$

Matrix and Random Vectors IV

- Result: Let A be a $n \times n$ square symmetric matrix. Then A has n pairs of eigenvalues and eigenvectors-namely,

$$\lambda_1, e_1; \lambda_2, e_2; \ldots; \lambda_n, e_n.$$

  The eigenvectors can be chosen to satisfy $1 = e_1' e_1 = ... = e_n' e_n$ and be mutually perpendicular. The eigenvectors are unique unless two or more eigenvalues are equal.

- Result: The spectral decomposition of a $n \times n$ symmetric matrix A is given by

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' \ldots + \lambda_n e_n e_n'.$$

- Example 2.10 (Page 61)

Matrix and Random Vectors V

- A square matrix A is said to be positve definite if

$$x'Ax > 0$$

for all vectors $x \neq 0$.

- Spectral Decomposition of square syemmetric

$$A = P \Lambda P',$$

where

$$P = [e_1 : e_2 : \ldots e_n] = \begin{bmatrix} e_{11} & e_{21} & \ldots & e_{n1} \\ e_{12} & e_{22} & \ldots & e_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \ldots & e_{nn} \end{bmatrix},$$

and

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

- Thus,
  - Inverse

$$A^{-1} = P\Lambda^{-1}P'.$$

  - Sqaure Root

$$A^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P'.$$

  - Factorization

$$A = A^{\frac{1}{2}}A^{\frac{1}{2}}.$$

- Cauchy-Schwarz Inequality: Let b and d any two $p \times 1$ vectors. Then
$$(b'd)^2 \leq (b'b)(d'd)$$
with equality iff $b = cd$ for some constant c.

Matrix inequalities and Maximization II

- Extended Cauchy-Schwarz Inequality: Let b, d be any two $p \times 1$ vectors and B be a positive definite matrix. Then

$$(b'd)^2 \leq (b'Bb)(d'B^{-1}d)$$

with equality iff $b = cB^{-1}d$ for some constant c.

Matrix inequalities and Maximization III

- Maximization Lemma: Let B be positive definite and d be a given vector. Then, for an arbitrary nonzero vector x,

$$\max_{x \neq 0} \frac{(x'd)^2}{x'Bx} = d'B^{-1}d$$

with the maximum attained when $x = cB^{-1}d$ for any constant $c \neq 0$.

Matrix inequalities and Maximization IV

- Maximization of Quadratic Forms for Points on the Unit Sphere: Let $B_{p\times p}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_p \geq 0$ and associated normalized eigenvectors $e_1, e_2, \ldots, e_p$. Then

  1. $$\max_{x\neq 0} \frac{x'Bx}{x'x} = \lambda_1 \text{ attained when } x = e_1$$

     and

  2. $$\min_{x\neq 0} \frac{x'Bx}{x'x} = \lambda_p \text{ attained when } x = e_p$$

  3. Moreover, for $k = 1, \ldots, p-1$

     $$\max_{x\perp e_1,\ldots,e_k} \frac{x'Bx}{x'x} = \lambda_{k+1} \text{ attained when } x = e_{k+1}.$$

- Sketch of proof:
  Let $B = P\Lambda P'$ and $y = P'x$, where $P = [e_1 : e_2 : \ldots : e_p]$ and $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$.
  1. Thus,

$$\frac{x'Bx}{x'x} = \frac{x'P\Lambda P'x}{x'PP'x} = \frac{y'\Lambda y}{y'y}$$

Hence,

$$\max_{x\neq 0} \frac{x'Bx}{x'x} \Leftrightarrow \max_{y\neq 0} \frac{y'\Lambda y}{y'y}$$

Now,

$$\frac{y'\Lambda y}{y'y} = \frac{\sum_{i=1}^{p} \lambda_i y_i^2}{\sum_{i=1}^{p} y_i^2} \leq \lambda_1 \frac{\sum_{i=1}^{p} y_i^2}{\sum_{i=1}^{p} y_i^2} = \lambda_1.$$

Also, the maximum is attained at $y = [1, 0, \ldots, 0]'$, equivalently at $x = Py = e_1$.

2. Similarly.
3. Note that

$$\mathbf{x} = P\mathbf{y} = y_1\mathbf{e_1} + \ldots + y_i\mathbf{e_i} + \cdots + y_p\mathbf{e_p}$$

and

$$\mathbf{e_i'x} = y_1\mathbf{e_i'e_1} + \ldots + y_i\mathbf{e_i'e_i} + \cdots + y_p\mathbf{e_i'e_p} = y_i$$

Hence,

$$\mathbf{x}\perp\mathbf{e}_1, \ldots, \mathbf{e}_k \Rightarrow y_i = 0 \forall i \leq k$$

Matrix inequalities and Maximization VII

Thus,

$$\max_{\mathbf{x}\perp\mathbf{e}_1,\dots,\mathbf{e}_k} \frac{\mathbf{x}'B\mathbf{x}}{\mathbf{x}'\mathbf{x}} \Leftrightarrow \max_{\mathbf{y}:y_i=0\forall i\leq k} \frac{\mathbf{y}'\Lambda\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

Therefore,

$$\frac{y'\Lambda y}{y'y} = \frac{\sum_{i=1}^{p}\lambda_i y_i^2}{\sum_{i=1}^{p} y_i^2} = \frac{\sum_{i=k+1}^{p}\lambda_i y_i^2}{\sum_{i=k+1}^{p} y_i^2} \leq \lambda_{k+1}\frac{\sum_{i=k+1}^{p} y_i^2}{\sum_{i=k+1}^{p} y_i^2} = \lambda_{k+1}.$$

Also, the maximum is attained at $y = [\underbrace{0,\dots,0}_{k},1,\dots,0]'$, equivalently at $x = Py = e_{k+1}$.

Random Vectors I

- Random vector: Vector of random variables

$$\underline{X} = \left[X_1, X_2, \ldots, X_p\right]'$$

- Mean vector

$$E(\underline{X}) = \left[\mu_1, \mu_2, \ldots, \mu_p\right]' = \underline{\mu}$$

- Covariance matrix

$$Cov(\underline{X}) = E(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} = \Sigma.$$

- Example 2.13 (Page 70)

- Correlation matrix

$$Cor(\underline{X}) = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}\sigma_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} = \rho.$$

Random Vectors III

- Standard deviation matrix

$$V^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{\sigma_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\sigma_{pp}} \end{bmatrix}.$$

- Relation between Σ and ρ through V.

$$\Sigma = V^{\frac{1}{2}} \rho V^{\frac{1}{2}}$$

and

$$\rho = V^{-\frac{1}{2}} \Sigma V^{-\frac{1}{2}}.$$

- Example 2.14 (Page 72)

Random Vectors IV

- Result: For any real constant vector $\underline{c} = [c_1, c_2, \ldots, c_p]'$, the linear combination $\underline{c}'\underline{X} = c_1X_1 + c_2X_2 + \ldots + c_pX_p$ has mean

$$E(\underline{c}'\underline{X}) = \underline{c}'\underline{\mu}$$

and variance

$$Var(\underline{c}'\underline{X}) = \underline{c}'\Sigma\underline{c}.$$

Random Vectors V

- Result: For any real matrix

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}$$

the linear combination $\underline{Z} = C\underline{X}$ has mean

$$\underline{\mu}_Z = E(\underline{Z}) = E(C\underline{X}) = C\underline{\mu}_X$$

and variance

$$\Sigma_Z = Cov(\underline{Z}) = Cov(C\underline{X}) = C\Sigma_X C'.$$

Random Vectors VI

- Result: For any two random vectors $\underline{X}_1$ and $\underline{X}_2$ of same order, let $\underline{Z} = \underline{X}_1 + \underline{X}_2$

$$
\begin{aligned}
\mu_Z &= E[\underline{X}_1 + \underline{X}_2] \\
&= E[\underline{X}_1] + E[\underline{X}_2] \\
&= \mu_1 + \mu_2.
\end{aligned}
$$

and

$$
\begin{aligned}
\Sigma_Z &= Var[\underline{X}_1 + \underline{X}_2] \\
&= Var[\underline{X}_1] + Var[\underline{X}_2] + Cov[\underline{X}_1, \underline{X}_2] + Cov[\underline{X}_2, \underline{X}_1] \\
&= \Sigma_{11} + \Sigma_{22} + \Sigma_{12} + \Sigma_{21}.
\end{aligned}
$$

Random Samples I

- Let $\mathbf{X}_{1\cdot}, \mathbf{X}_{2\cdot}, \ldots, \mathbf{X}_{n\cdot}$ be n samples drawn from a random distribution. Then the sample mean $\bar{\mathbf{X}}$ is calculated as

$$\begin{aligned}
\bar{\mathbf{X}} &= \frac{1}{n}[\mathbf{X}_{1\cdot} + \mathbf{X}_{2\cdot} + \cdots + \mathbf{X}_{n\cdot}]' \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} X_{i1}, \sum_{i=1}^{n} X_{i2}, \ldots, \sum_{i=1}^{n} X_{ip}\right]' \\
&= [\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p]'
\end{aligned}$$

Random Samples II

and the sample variance is calculated as

$$S_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_{i.} - \bar{\mathbf{X}})(\mathbf{X}_{i.} - \bar{\mathbf{X}})'$$

$$= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} X_{i1} - \bar{X}_1 \\ X_{i2} - \bar{X}_2 \\ \vdots \\ X_{ip} - \bar{X}_p \end{bmatrix}\begin{bmatrix} X_{i1} - \bar{X}_1 & X_{i2} - \bar{X}_2 & \cdots & X_{ip} - \bar{X}_p \end{bmatrix}$$

$$= \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n}(X_{i1}-\bar{X}_1)^2 & \sum_{i=1}^{n}(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2) & \cdots & \sum_{i=1}^{n}(X_{i1}-\bar{X}_1)(X_{ip}-\bar{X}_p) \\ \sum_{i=1}^{n}(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2) & \sum_{i=1}^{n}(X_{i2}-\bar{X}_2)^2 & \cdots & \sum_{i=1}^{n}(X_{i2}-\bar{X}_2)(X_{ip}-\bar{X}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}(X_{i1}-\bar{X}_1)(X_{ip}-\bar{X}_p) & \sum_{i=1}^{n}(X_{i2}-\bar{X}_2)(X_{ip}-\bar{X}_p) & \cdots & \sum_{i=1}^{n}(X_{ip}-\bar{X}_p)^2 \end{bmatrix}$$

- Example 1.2 (Page 10)

Random Samples III

- Let $\mathbf{X}_{1\cdot}, \mathbf{X}_{2\cdot}, \ldots, \mathbf{X}_{n\cdot}$ be random samples from a joint distribution that has mean vector $\underline{\mu}$ and covariance matrix Σ. Then for the sample mean $\bar{\mathbf{X}}$,

$$E(\bar{\mathbf{X}}) = \underline{\mu} \text{ and } Cov(\bar{\mathbf{X}}) = \frac{1}{n}\Sigma$$

and for the sample variance S_n,

$$E(S_n) = \frac{n-1}{n}\Sigma$$

Random Samples IV

- Therefore an unbiased estimator of Σ is

$$
\begin{aligned}
S &= \left(\frac{n}{n-1}\right) S_n \\
&= \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_{i\cdot} - \bar{\mathbf{X}})(\mathbf{X}_{i\cdot} - \bar{\mathbf{X}})' \\
&= \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{12} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1p} & S_{2p} & \dots & S_{pp} \end{bmatrix}
\end{aligned}
$$

Generalized Sample Variance I

- Determinant of S is called as generalized sample variance.
- One can show that for a $p-$variate data set

$$\text{Generalized Sample Variance} = |S| = (n-1)^{-p}(\textit{hyper volume})^2$$

  by induction.
  - Geometrical interpretation for bivariate data:
    Example 3.7 (Page 124)
- For highly correlated data generalized sample variance will be smaller.

Statistical Distance I

- Statistical distance (d) between any two sample points

$$P = X_{i\cdot} \text{ and } Q = X_{j\cdot}$$

in a sample set $\{X_{1\cdot}, X_{2\cdot}, \ldots, X_{n\cdot}\}$ is defined as

$$d^2(P, Q) = (X_{i\cdot} - X_{j\cdot})' S^{-1} (X_{i\cdot} - X_{j\cdot})$$

- Figure 1.25 (Page 37)