Vectors and Matrices

Vector

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Inner product

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots x_n y_n$$

Vector norm (length)

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{x \cdot x}$$

Inner product: Geometric Interpretation

$$\cos\theta = \frac{x \cdot y}{\|x\| \cdot \|y\|}$$

x and y are **orthogonal** if $x \cdot y = 0$

Linear dependence

Vectors $x_1, \ldots x_m$ are **linearly dependent** if there exist scalars $a_1, \ldots a_m$ such that at least one $a_j \neq 0$ and

$$a_1 x_1 + \cdots + a_m x_m = 0$$

Vector space

- Basis vectors: $x_1, \dots x_m$
- $\mathcal{X} = \mathrm{span}(x_1, \dots, x_m)$
  $= \{x : x = a_1 x_1 + a_2 x_2 + \cdots + a_m x_m, a_j \in \mathbb{R}\}$

Matrix

$$A_{n\times m} = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nm} \end{pmatrix}$$

Matrix multiplication

$$
\begin{aligned}
A_{n\times m}B_{m\times k} &= \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nm} \end{pmatrix} \cdot \begin{pmatrix} B_{11} & \cdots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mk} \end{pmatrix} \\
&= \begin{pmatrix} \cdots & \cdots & \cdots \\ \cdots & \sum_{\ell=1}^{m} A_{i\ell}B_{\ell j} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}_{n\times k}
\end{aligned}
$$

Matrix transpose

$$A^T_{n\times m} = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \ddots & \ddots & \vdots \\ A_{1m} & A_{2m} & \cdots & A_{nm} \end{pmatrix}_{m\times n}$$

$$(AB)^T = B^T A^T$$

- Matrix **inverse** (of a square matrix $A_{n \times n}$)
  - Defined by $A^{-1}A = AA^{-1} = I$
  - If the columns (rows) of A are **linearly independent**, then A is **invertible**; otherwise, A is **singular**, and **determinant** $\det(A) = 0$.
- Properties

$$
\begin{aligned}
&(AB)^{-1} = B^{-1}A^{-1} \\
&(A^{-1})^{T} = (A^{T})^{-1} = A^{-T} \\
&(AB)^{-T} = A^{-T}B^{-T}
\end{aligned}
$$

Eigendecomposition

- **Eigenvectors and eigenvalues**: there exists a vector $u \neq 0$

$$A_{n \times n} u_{n \times 1} = \lambda u_{n \times 1}$$

- There are exactly n eigenvalues and eigenvectors (not all eigenvalues are distinct and/or real)
- If A is **symmetric** ($A^T = A$), all eigenvalues are real
- A is singular iff at least one of the $\lambda = 0$
- $\lambda \neq 0$ an eigenvalue of $A \Leftrightarrow 1/\lambda$ an eigenvalue of A^{-1}
- A is **positive definite** ($x^T A x > 0$ for all $x \neq 0$)$\Leftrightarrow$ all eigenvalues $\lambda > 0$. Positive semi-definite: all $\lambda \geq 0$.
- $\det(A) = \lambda_1 \times \cdots \times \lambda_n$

Derivatives of vector-valued functions

- Argument $x = (x_1, \ldots x_m)$
- Function $y_{n\times 1} = f(x_{m\times 1})$;
- Derivative $D_{m\times n} = \partial y / \partial x$ is defined by

$$D_{ij} = \frac{\partial y_j}{\partial x_i}$$

- $\partial(A_{n\times m} x_{m\times 1})/\partial x = A^T$
- $\partial(x^T B_{m\times m} x)/\partial x = (B + B^T)x$

Probability and Statistics

- **Continuous** random variables
- Probability density function (**p.d.f.**)

$$f(z) = \lim_{\Delta z \downarrow 0} \frac{Pr(z \leq Z \leq z + \Delta z)}{\Delta z}$$

Example: Normal distribution

$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(z-\mu)^2/2\sigma^2}, \quad -\infty < z < \infty$$

Standard normal: $\mu = 0, \sigma = 1$ (**parameters**)

- Cumulative distribution function (**c.d.f.**)

$$
\begin{aligned}
F(z) &= P(Z \le z) = \int_{-\infty}^{z} f(z')dz' \\
f(z) &= \frac{dF(z)}{dz}
\end{aligned}
$$

- Properties

$$f(z) \ge 0$$

$$0 \le F(z) \le 1$$

$$P(a \le Z \le b) = \int_{a}^{b} f(z)dz = F(b) - F(a)$$

$$\int_{-\infty}^{\infty} f(z)dz = 1$$

- **Quantile**: the $100\alpha\%$ quantile q_α satisfies

$$\int_{-\infty}^{q_\alpha} f(z)dz = \alpha$$

Example: $\pm 2\sigma$ for normal covers about 95%

- **Median**: the 50% quantile
- For symmetric distributions, median = mean.

- **Mean and variance**

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} zf(z)dz \\ Var(Z) &= \int_{-\infty}^{\infty} (z - E(Z))^2 f(z)dz \end{aligned}$$

- **Covariance and correlation**

$$
\begin{aligned}
Cov(Z_1, Z_2) &= E\left((Z_1 - E(Z_1))(Z_2 - E(Z_2))\right) \\
Cov(Z_1, Z_2) &= Cov(Z_2, Z_1) \\
Cor(Z_1, Z_2) &= \frac{Cov(Z_1, Z_2)}{\sqrt{Var(Z_1)}\sqrt{Var(Z_2)}}
\end{aligned}
$$

If Z_1 and Z_2 are **independent**, $Cov(Z_1, Z_2) = 0$

- Properties

$$
\begin{aligned}
E(aZ + b) &= aE(Z) + b \\
Var(aZ + b) &= a^2 Var(Z) \\
E(Z_1 + \cdots + Z_m) &= E(Z_1) + \cdots + E(Z_m) \\
Var(Z_1 + \cdots + Z_m) &= Var(Z_1) + \cdots + Var(Z_m) \\
&\quad + 2\sum_{j<j'} Cov(Z_j, Z_{j'})
\end{aligned}
$$

- Vector of r.v.s: $Z = (Z_1, \ldots Z_m)$

$$\begin{aligned} E(Z) &= (E(Z_1), \ldots E(Z_m)) \\ Cov(Z) &= \begin{pmatrix} Var(Z_1) & \cdots & \cdots & Cov(Z_1, Z_m) \\ Cov(Z_2, Z_1) & Var(Z_2) & \cdots & Cov(Z_2, Z_m) \\ \vdots & \ddots & \ddots & \vdots \\ Cov(Z_m, Z_1) & \cdots & \cdots & Var(Z_m) \end{pmatrix} \end{aligned}$$

If Z_j's are independent, $Cov(Z) = diag(Var(Z_j))$.

$$\begin{aligned} E(A_{n\times m}Z) &= A \cdot E(Z) \\ Cov(A_{n\times m}Z) &= A \cdot Cov(Z) \cdot A^T \end{aligned}$$

Multivariate Normal Distribution

- $(X_1, \ldots, X_p)$ is $p \times 1$ random vector, with mean $\mu_{p\times 1}$, variance $\Sigma_{p\times p}$
- The joint density at $x = (x_1, \ldots, x_p)$ is

$$f(x; \mu, \Sigma) = (2\pi \det(\Sigma))^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

- Write this as $X \sim N_p(\mu, \Sigma)$
- Each coordinate X_i has normal distribution $N(\mu_i, \Sigma_{ii})$
- $\Sigma_{ij} = 0$ if **and only if** X_i, X_j independent