Review (Probability & Linear Algebra)

CE-725 : Statistical Pattern Recognition  
Sharif University of Technology  
Spring 2013  

M. Soleymani
Outline

- Axioms of probability theory
- Conditional probability, Joint probability, Bayes rule
- Discrete and continuous random variables
  - Probability mass and density functions
  - Expected value, variance, standard deviation
- Expectation for two variables
  - covariance, correlation
- Some probability distributions
  - Gaussian distribution
- Linear Algebra
Basic Probability Elements

- Sample space ($\Omega$): set of all possible outcomes (or worlds)
  - Outcomes are assumed to be mutually exclusive.

- An event $A$ is a certain set of outcomes (i.e., subset of $\Omega$).

- A random variable is a function defined over the sample space
  - Gender: $Students \rightarrow \{\text{male, female}\}$
  - Height: $Students \rightarrow \mathbb{R}^+$
Probability Space

- A probability space is defined as a triple \((\Omega, F, P)\):
  - A sample space \(\Omega \neq \emptyset\) that contains the set of all possible outcomes (outcomes also called states of nature)
  - A set \(F\) whose elements are called events. The events are subsets of \(\Omega\). \(F\) should be a “Borel Field”.
  - \(P\) represents the probability measure that assigns probabilities to events.
Probability and Logic

- Probability theory can be viewed as a generalization of propositional logic
  - Probabilistic logic

- $P(a)$:
  - $a$ is a propositional logic sentence
  - Belief of agent in $a$ is no longer restricted to true, false, unknown
  - $P(a)$ can range from 0 to 1
Probability Axioms (Kolomogrov)

- Axioms define a reasonable theory of uncertainty

- Kolomogrov’s probability axioms (propositional notation)
  - $P(a) \geq 0 \quad (\forall a \in S)$
  - If $T$ is a tautology $P(T) = 1$
  - $P(a \lor b) = P(a) + P(b) - P(a \land b) \quad (\forall a, b \in S)$
Random Variables

- Random variables: Variables in probability theory

- Domain of random variables: **Boolean, discrete** or **continuous**

- **Probability distribution**: the function describing probabilities of possible values of a random variable
  - \( P(Dice = 1) = \frac{1}{6}, P(Dice = 2) = \frac{1}{6}, \ldots \)
Random Variables

- Random variable is a function that maps every outcome in $\Omega$ to a real (complex) number.
  - To define probabilities easily as functions defined on (real) numbers.
  - Expectation, variance, …
Probabilistic Inference

- **Joint probability distribution**
  - Can specify probability of every atomic event
  - We can find every probabilities from it (by summing over atomic events).

- **Prior and posterior probabilities**
  - belief in the absence or presence of evidences

- **Bayes’ rule**
  - used when it is difficult to compute $P(a|b)$ but we have information about $P(b|a)$

- **Independence**
  - new evidence may be irrelevant
Joint Probability Distribution

- Probability of all combinations of the values for a set of random variables.

- If two or more random variables are considered together, they can be described in terms of their joint probability.

- Example: Joint probability of features
  - \( p(x_1, x_2, ..., x_d) \)
Prior and Posterior Probabilities

- **Prior** or unconditional probabilities of propositions: belief in the absence of any other evidence
  - e.g., \( P(\text{Day} = \text{Sunday}) = 0.14 \)
  \[ P(\text{Weather} = \text{sunny}) = 0.6 \]

- **Posterior** or conditional probabilities: belief in the presence of evidences
  - \( P(a|b) \) is the probability of \( a \) given that \( b \) is true
  - e.g., \( P(\text{Weather} = \text{rainy} | \text{Season} = \text{Spring}) \)
Conditional Probability

\[ P(a|b) = \frac{P(a \land b)}{P(b)} \quad \text{if } P(b) > 0 \]

- **Product rule** (an alternative formulation):
  \[ P(a \land b) = P(a|b) \cdot P(b) = P(b|a) \cdot P(a) \]

- \( P(a|b) \) obeys the same rules as probabilities
  - E.g., \( P(a|b) + P(\sim a|b) = 1 \)
Conditional Probability

- For statistically dependent variables, knowing the value of one variable may allow us to better estimate the other.

- All probabilities in effect are conditional probabilities
  - E.g., \( P(a) = P(a \mid \text{our background knowledge}) \)

Renormalize the probability of events jointly occur with b
Conditional Probability: Example

- Rolling a fair dice
  - \(a\) : the outcome is an even number
  - \(b\) : the outcome is a prime number

\[
P(a|b) = \frac{P(a \land b)}{P(b)} = \frac{1/6}{1/2} = \frac{1}{3}
\]
Chain rule

- Chain rule is derived by successive application of product rule:

\[
P(X_1, \ldots, X_n) = \prod_{i=2}^{n} P(X_i | X_1, \ldots, X_{i-1})
\]
Law of Total Probability

- $B_1, \ldots, B_N$ mutually exclusive and $\bigcup_{i=1}^{N} B_i = \Omega$
- $A$ is an event (subset of $\Omega$)

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_N)$$

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_N)P(B_N)$$
Independence

- Propositions $a$ and $b$ are independent iff

\[
\begin{align*}
P(a|b) &= P(a) \\
P(b|a) &= P(b) \\
P(a, b) &= P(a)P(b)
\end{align*}
\]

- Knowing $a$ tells us nothing about $b$ (and vice versa)
Bayes Rule

- **Bayes rule:** \( P(a|b) = \frac{P(b|a)P(a)}{P(b)} \)
  - Obtained from product rule: \( P(a \land b) = P(a|b)P(b) = P(b|a)P(a) \)

- In some problems, it may be difficult to compute \( P(a|b) \) directly, yet we might have information about \( P(b|a) \).

  \[
P(Cause|Effect) = P(Effect|Cause) \cdot P(Cause) / P(Effect)
\]
Bayes Rule

Often it would be useful to derive the rule a bit further:

\[ P(a|b) = \frac{P(b|a)P(a)}{P(b|a)P(a) + P(b|\neg a)P(\neg a)} \]

\[ = \frac{P(a \land b)}{P(b \land \neg a) + P(b \land a)} \]
Bayes Rule: Example

- **Meningitis**($M$) & **Stiff neck** ($S$)
  - $P(m) = \frac{1}{5000}$
  - $P(s) = 0.01$
  - $P(s|m) = 0.7$

- $P(m|s) =$?

  $$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.7 \times 0.0002}{0.01} = 0.0014$$
Bayes Rule: General Form

\[ P(B_j | A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A \cap B_j)}{\sum_{i=1}^{N} P(A \cap B_i)} \]
Calculating Probabilities from Joint Distribution

- **Useful techniques**
  - Marginalization
    - \( P(Y) = \sum_{z \in Z} P(Y, z) \)
  - Conditioning
    - \( P(Y) = \sum_{z \in Z} P(Y|z)P(z) \)
  - Normalization
    - \( P(X|e) = \alpha P(X, e) \)
Probability Mass Function (PMF)

- Probability Mass Function (PMF) shows the probability for each value of a discrete random variable.
  - Each impulse magnitude is equal to the probability of the corresponding outcome.

- Example: PMF of a fair dice.

\[
P(X = x) \geq 0
\]
\[
\sum_{x \in X} P(X = x) = 1
\]
Probability Density Function (PDF)

- Probability Density Function (PDF) is defined for continuous random variables
  - The probability of \( x \in \left( x_0 - \frac{\epsilon}{2}, x_0 + \frac{\epsilon}{2} \right) \) is \( p_X(x_0) \times \epsilon \)

\[
p_X(x_0) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} P \left( x_0 - \frac{\epsilon}{2} \leq X \leq x_0 + \frac{\epsilon}{2} \right)
\]

- \( p_X(x) \geq 0 \)
- \( \int p_X(x)dx = 1 \)
Cumulative Distribution Function (CDF)

- Defined as the integration of PDF
- Similarly defined on discrete variables (summation instead of integration)

\[
CDF(x) = F_X(x) = P(X \leq x)
\]

\[
F_X(x) = \int_{-\infty}^{x} p_X(\tau) d\tau
\]

Non-decreasing
Right Continuous
\[
F_X(-\infty) = 0
\]
\[
F_X(\infty) = 1
\]

\[
\frac{dF_X(x)}{dx} = f_X(x)
\]

\[
P(u < x \leq v) = F_X(v) - F_X(u)
\]
Distribution Statistics

- Basic descriptors of spatial distributions:
  - Mean Value
  - Variance & Standard Deviation
  - Moments
  - Covariance & Correlation
Expected Value

- Expected (or mean) value: weighted average of all possible values of the random variable

- Expectation of a discrete random variable $X$:
  \[ E[X] = \sum_{x} xP(x) \]

- Expectation of a function of a discrete random variable $X$:
  \[ E[g(X)] = \sum_{x} g(x)P(x) \]

- Expected value of a continuous random variable $X$:
  \[ E[X] = \int_{-\infty}^{\infty} xp_X(x)dx \]

- Expectation of a function of a continuous random variable $X$:
  \[ E[g(X)] = \int_{-\infty}^{\infty} g(x)p_X(x)dx \]
Variance

- Variance: a measure of how far values of a random variable are spread out around its expected value

\[ Var(X) = E \left[ (X - \mu_X)^2 \right] \]

\[ = E \left[ X^2 \right] - \mu_X^2 \]

\[ \mu_X = E[ X ] \]

- Standard deviation: square root of variance

\[ \sigma_X = \sqrt{\text{var}(X)} \]
Moments

- Moments
  - $n^{th}$ order moment of a random variable $X$:
    \[ M_n = E[X^n] \]
  - Normalized $n^{th}$ order moment:
    \[ E[(X - \mu_X)^n] \]

- The first order moment is the mean value.
- The second order moment is the variance added by the square of the mean.
Correlation & Covariance

- **Correlation**

\[
Crr(X, Y) = E[XY] \Rightarrow \sum_x \sum_y xyP(x, y)
\]

- **Covariance** is the correlation of mean removed variables:

\[
Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y
\]
Covariance: Example

\[ \text{Cov}(X, Y) = 0 \]

\[ \text{Cov}(X, Y) = 0.9 \]
Covariance Properties

- The covariance value shows the tendency of the pair of RVs to increase together
  - $C_{XY} > 0$: $X$ and $Y$ tend to increase together
  - $C_{XY} < 0$: $X$ tends to decrease when $Y$ increases
  - $C_{XY} = 0$: no linear correlation between $X$ and $Y$
Orthogonal, Uncorrelated & Independent RVs

- **Orthogonal** random variables \((E[XY] = 0)\)
  - \(Crr(X, Y) = 0\)

- **Uncorrelated** random variables \((E[(X - \mu_X)(Y - \mu_Y)] = 0)\)
  - \(Cov(X, Y) = 0\)

- **Independent** random variables \(\Rightarrow Cov(X, Y) = 0\)
  - \(Cov(X, Y) = 0 \Leftrightarrow \text{Independent} \text{ random variables}\)
Pearson’s Product Moment Correlation

\[ \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \]

- Defined only if both \( \sigma_X \) and \( \sigma_Y \) are finite and nonzero.
- \(|\rho_{XY}|\) shows the degree of linear dependence between \( X \) and \( Y \).
- \(-1 \leq \rho_{XY} \leq 1\)
  - \( E[(X - \mu_X)(Y - \mu_Y)]^2 \leq E[(X - \mu_X)^2]E[(Y - \mu_Y)^2] \) according to Cauchy-Schwarz inequality (\( C_{XY} \leq \sigma_X \sigma_Y \))
  - \( \rho_{XY} = 1 \) shows a perfect positive linear relationship and \( \rho_{XY} = -1 \) shows a perfect negative linear relationship
Pearson’s Correlation: Examples

\[ \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \]
Covariance Matrix

- If \( x \) is a vector of random variables (\( d \)-dim random vector):
  - Covariance matrix indicates the tendency of each pair of RVs to vary together

\[
\Sigma = E[(x - \mu_x)(x - \mu_x)^T] = \begin{bmatrix}
E((x_1 - \mu_1)(x_1 - \mu_1)) & \cdots & E((x_1 - \mu_1)(x_d - \mu_d)) \\
\vdots & \ddots & \vdots \\
E((x_d - \mu_d)(x_1 - \mu_1)) & \cdots & E((x_d - \mu_d)(x_d - \mu_d))
\end{bmatrix}
\]

\[
\mu_x = \begin{bmatrix}
\mu_1 \\
\vdots \\
\mu_d
\end{bmatrix} = \begin{bmatrix}
E(x_1) \\
\vdots \\
E(x_d)
\end{bmatrix}
\]
Covariance Matrix: Two Variables

\[ \Sigma = C = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \]

\[ \sigma_{12} = \sigma_{21} = \text{Cov}(X, Y) \]

\[ \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]

\[ \Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \]
Covariance Matrix

- $\sigma_{ij}$ shows the covariance of $X_i$ and $X_j$:

$$\sigma_{ij} = \text{Cov}(X_i, X_j) = E \left[ (X_i - \mu_i)(X_j - \mu_j) \right]$$

$$\Sigma = \begin{bmatrix}
\sigma_{1}^2 & \sigma_{12} & \cdots & \sigma_{1d} \\
\sigma_{21} & \sigma_{2}^2 & \cdots & \sigma_{2d} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{d}^2 
\end{bmatrix}$$
Sums of Random Variables

\[ Z = X + Y \]

- **Mean:** \( \mu_Z = \mu_X + \mu_Y \)

- **Variance:** \( Var(Z) = Var(X) + Var(Y) + 2Cov(X, Y) \)
  - If \( X, Y \) independent: \( Var(Z) = Var(X) + Var(Y) \)

- **Distribution:**
  \[
  p_Z(z) = \int p_{X,Y}(x, z-x) dx \\
  \text{independence} \\
  = p_X(x) * p_Y(y) = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x) dx
  \]
Some Famous Probability Mass and Density Functions

- **Uniform**

  \[ p(x) = \frac{1}{b-a} \left( U(a) - U(b) \right) \]

- **Gaussian (Normal)**

  \[ p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} = N(\mu, \sigma) \]
Some Famous Probability Mass and Density Functions

- **Binomial**

  \[ P(X = k) = \binom{n}{k} (1 - p)^{n-k} p^k \]

- **Exponential**

  \[ p(x) = \lambda e^{-\lambda x} U(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \]
Gaussian (Normal) Distribution

- It is widely used to model the distribution of continuous variables

- Standard Normal distribution: $\mu = 0, \sigma = 1$

68% within $[\mu - \sigma, \mu + \sigma]$

95% within $[\mu - 2\sigma, \mu + 2\sigma]$
Multivariate Gaussian Distribution

- $x$ is a vector of $d$ Gaussian variables

$$p(x) \sim N(\mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

$$\mu = E[x] = [E(x_1), \ldots, E(x_d)]^T$$

$$\Sigma = E[(x-\mu)(x-\mu)^T]$$

- Mahalanobis Distance:

$$r^2 = (x-\mu)^T \Sigma^{-1}(x-\mu)$$
Multivariate Gaussian Distribution

- The covariance matrix is always symmetric and positive semi-definite.

- Multivariate Gaussian is completely specified by \( d + d(d + 1)/2 \) parameters.

- Special cases of \( \Sigma \):
  - \( \Sigma = \sigma^2 I \): Independent random variables with the same variance (circularly symmetric Gaussian).
  - Digonal matrix \( \Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_d^2 \end{bmatrix} \): Independent random variables with different variances.
Multivariate Gaussian Distribution
Level Surfaces

- The Gaussian distribution will be constant on surfaces in $x$-space for which:
  \[(x - \mu)^T \Sigma^{-1} (x - \mu) = C\]  
  Hyper-ellipsoid

- Principal axes of the hyper-ellipsoids are the eigenvectors of $\Sigma$.

- Bivariate Gaussian: Curves of constant density are ellipses.
Bivariate Gaussian distribution

- $\lambda_1$ and $\lambda_2$ are the eigenvalues of $\Sigma$ ($\lambda_1 \geq \lambda_2$) and $\nu_1$ and $\nu_2$ are the corresponding eigenvectors

$$\frac{l_1}{l_2} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}$$
Linear Transformations on Multivariate Gaussian

\[ A_w : \text{Whitening transform} \]

\[ A_w = V \Lambda^{-\frac{1}{2}} \]
Some attracting properties of the Gaussian distribution:

- Marginal and conditional distributions of a Gaussian are also Gaussian.
- After a linear transformation, a Gaussian distribution is again Gaussian.
  - There exists a linear transformation that diagonalizes the covariance matrix (whitening transform).
    - It converts the multivariate normal distribution into a spherical one.
- Gaussian is a distribution that maximizes the entropy.
- Gaussian is stable and infinitely divisible.
  - Central Limit Theorem
- Some distributions can be estimated by Gaussian distribution when their parameter value is sufficiently large (e.g., Binomial).
Central Limit Theorem (under mild conditions)

- Suppose i.i.d. (Independent Identically Distributed) RVs $X_i (i = 1, \ldots, N)$ with finite variances

- Let $S_N = \sum_{i=1}^{N} X_i$ be the sum of these RVs

- Distribution of $S_N$ converges to a normal distribution as $N$ increases, regardless to the distribution of the RVs.

- Example:

  $X_i \sim \text{uniform}, i = 1, \ldots, N$

  $$S_N = \frac{1}{N} \sum_{i=1}^{N} X_i$$

  ![Histograms for different values of N](image)
Linear Algebra: Basic Definitions

- **Matrix A:**
  \[
  A = [a_{ij}]_{m \times n} = \begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
  \end{bmatrix}
  \]

- **Matrix Transpose**
  \[
  B = A^T \iff b_{ij} = a_{ji}, \quad 1 \leq i \leq n, 1 \leq j \leq m
  \]

- **Symmetric matrix A = A^T**

- **Vector a**
  \[
  a = \begin{bmatrix}
  a_1 \\
  \vdots \\
  a_n
  \end{bmatrix}, \quad a^T = [a_1, \ldots, a_n]
  \]
Linear Mapping

- Linear function
  - \( f(x + y) = f(x) + f(y) \quad \forall x, y \in V \)
  - \( f(ax) = af(x) \quad \forall x \in V, a \in F \)

- A linear function: \( f(x) = w_1x_1 + \cdots + w_dx_d = w^T x \)
- In general, a matrix \( W_{m \times d} = [w_1 \cdots w_m]^T \) can be used to denote a map \( f: \mathbb{R}^d \to \mathbb{R}^m \) where \( f_i(x) = w_{i1}x_1 + \cdots + w_{id}x_d = w_i^T x \)
Linear Algebra: Basic Definitions

- Inner (dot) product

\[ \langle a, b \rangle = a^T b = \sum_{i=1}^{n} a_i b_i \]

- Matrix multiplication

\[ A = [a_{ij}]_{m \times p} \quad B = [b_{ij}]_{p \times n} \]

\[ AB = C = [c_{ij}]_{m \times n} \quad c_{ij} = \langle A_i^T, B_j \rangle \]

- i-th row of A
- j-th column of B
Inner Product

- Inner (dot) product
  \[ \mathbf{a}^T \mathbf{b} = \sum_{i=1}^{n} a_i b_i \]

- Length (Euclidean norm) of a vector
  - \( \mathbf{a} \) is normalized iff \( ||\mathbf{a}|| = 1 \)
  \[ ||\mathbf{a}|| = \sqrt{\mathbf{a}^T \mathbf{a}} = \sqrt{\sum_{i=1}^{n} a_i^2} \]

- Angle between vectors
  \[ \cos \theta = \frac{\mathbf{a}^T \mathbf{b}}{||\mathbf{a}|| ||\mathbf{b}||} \]

- Orthogonal vectors \( \mathbf{a} \) and \( \mathbf{b} \):
  \[ \mathbf{a}^T \mathbf{b} = 0 \]

- Orthonormal set of vectors \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \):
  \[ \forall i, j \quad \mathbf{a}_i^T \mathbf{a}_j = \begin{cases} 1 & i = j \\ 0 & \text{o.w.} \end{cases} \]
Linear Independence

- A set of vectors is *linearly independent* if no vector is a linear combination of other vectors.

\[ c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \ldots + c_k \mathbf{v}_k = 0 \Rightarrow \]

\[ c_1 = c_2 = \ldots = c_k = 0 \]
Matrix Determinant and Trace

- **Determinant**
  - $\det(AB) = \det(A) \times \det(B)$
  - $\det(A) = \sum_{j=1}^{n} a_{ij} A_{ij} ; \quad i = 1, \ldots, n$
  - $A_{ij} = (-1)^{i+j} \det(M_{ij})$

- **Trace**
  - $tr[A] = \sum_{j=1}^{n} a_{jj}$
Matrix Inversion

- **Inverse of** $A_{n \times n}$:
  
  $$AB = BA = I_n \Leftrightarrow B = A^{-1}$$

- $A^{-1}$ exists iff $\det(A) \neq 0$ (A is nonsingular)
  
  - Singular: $\det(A) = 0$
  - ill-conditioned: A is nonsingular but close to being singular

- **Pseudo-inverse** for a non square matrix $A^\# = (A^T A)^{-1} A^T$
  
  - $A^T A$ is not singular
  - $A^\# A = I$
Matrix Rank

- $\text{rank}(A)$: maximum number of linearly independent columns or rows of $A$.

- $A_{m \times n}$: $\text{rank}(A) \leq \min(m, n)$

- **Full rank** $A_{n \times n}$: $\text{rank}(A) = n$ iff $A$ is nonsingular ($\det(A) \neq 0$)
Eigenvectors and Eigenvalues

\[ Av = \lambda v \]

- **Characteristic equation:** \( \det(A - \lambda I_n) = 0 \)
  - \( n \)-th order polynomial, with \( n \) roots

\[ \text{tr}(A) = \sum_{j=1}^{n} \lambda_j \]

\[ \det(A) = |A| = \prod_{j=1}^{n} \lambda_j \]
Eigenvalues and Eigenvalues: Symmetric Matrix

- For a symmetric matrix, the eigenvectors corresponding to distinct eigenvalues are orthogonal.

- These eigenvectors can be used to form an orthonormal set ($\forall i \neq j \quad \mathbf{v}_i^T \mathbf{v}_j = 0$ and $\|\mathbf{v}_i\| = 1$).
Eigen Decomposition: Symmetric Matrix

\[ V = [v_1 \ldots v_N], \Lambda = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_N \end{bmatrix} \]

\[ AV = V \Lambda \Rightarrow AVV^T = V \Lambda V^T \overset{VV^T = I}{\rightarrow} A = V \Lambda V^T \]

- Eigen decomposition of a symmetric matrix: \( A = V \Lambda V^T \)
Positive definite matrix

- Symmetric $A_{n \times n}$ is positive definite:
  \[ \forall x \in \mathbb{R}^n \Rightarrow x^T A x > 0 \]

- Eigen values of a positive define matrix are positive:
  \[ \forall i, \lambda_i > 0 \]
Simple Vector Derivatives

\[
\frac{\partial x^T A x}{\partial x} = (A + A^T)x
\]

\[
\frac{\partial b^T x}{\partial x} = b
\]