LECTURE 3: Review of Linear Algebra

- Vector and matrix notation
- Vectors
- Matrices
- Vector spaces
- Linear transformations
- Eigenvalues and eigenvectors

Vector and matrix notation

- **A d-dimensional (column) vector x and its transpose are written as:**

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \text{ and } x^T = \begin{bmatrix} x_1 \, x_1 \cdots x_d \end{bmatrix}$$

- **An n×d (rectangular) matrix and its transpose are written as:**

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1d} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & & a_{nd} \end{bmatrix} \text{ and } A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ a_{13} & a_{23} & \cdots & a_{n3} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1d} & a_{2d} & & a_{nd} \end{bmatrix}$$

- **The product of two matrices is**

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1d} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2d} \\ \vdots & \vdots & \vdots & \ddots & \\ a_{m1} & a_{m2} & a_{m3} & & a_{md} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ b_{31} & b_{32} & \cdots & b_{3n} \\ \vdots & \vdots & \ddots & \\ b_{d1} & b_{d2} & & b_{dn} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2n} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \\ c_{m1} & c_{m2} & c_{m3} & & c_{mn} \end{bmatrix} \text{ where } c_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj}$$

Vectors

- **The <u>inner product</u> (a.k.a. *dot product or scalar product*) of two vectors is defined by**

$$\langle x,y\rangle = x^T y = y^T x = \sum_{k=1}^{d} x_k y_k$$

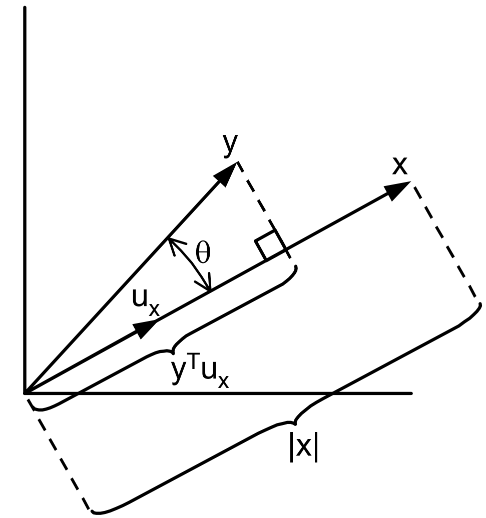
- **The <u>magnitude</u> of a vector is**

$$|x| = \sqrt{x^T x} = \left[\sum_{k=1}^{d} x_k x_k\right]^{1/2}$$

- **The <u>orthogonal projection</u> of vector y onto vector x is**

$$\langle y^T u_x\rangle u_x$$

  - where vector u_x has unit magnitude and the same direction as x



- **The <u>angle</u> between vectors x and y is**

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

- **Two vectors x and y are said to be**
  - <u>orthogonal</u> if $x^T y=0$
  - <u>orthonormal</u> if $x^T y=0$ <u>and</u> $|x|=|y|=1$
- **A set of vectors x_1, x_2, …, x_n are said to be <u>linearly dependent</u> if there exists a set of coefficients a_1, a_2, …, a_n (at least one different than zero) such that**

$$a_1x_1 + a_2x_2 + \cdots a_nx_n = 0$$

- **Alternatively, a set of vectors x_1, x_2, …, x_n are said to be <u>linearly independent</u> if**

$$a_1x_1 + a_2x_2 + \cdots a_nx_n = 0 \Rightarrow a_k = 0 \;\; \forall k$$

Matrices

- **The <u>determinant</u> of a square matrix $A_{d \times d}$ is**

$$|A| = \sum_{k=1}^{d} a_{ik} |A_{ik}| (-1)^{k+i}$$

  - where A_{ik} is the <u>minor</u> matrix formed by removing the i^{th} row and the k^{th} column of A
  - NOTE: the determinant of a square matrix and its transpose is the same: $|A|=|A^T|$
- **The <u>trace</u> of a square matrix $A_{d \times d}$ is the sum of its diagonal elements**

$$tr(A) = \sum_{k=1}^{d} a_{kk}$$

- **The <u>rank</u> of a matrix is the number of linearly independent rows (or columns)**
- **A square matrix is said to be <u>non-singular</u> if and only if its rank equals the number of rows (or columns)**
  - A non-singular matrix has a non-zero determinant
- **A square matrix is said to be <u>orthonormal</u> if $AA^T=A^TA=I$ (more on this later)**
- **For a square matrix A**
  - if $x^TAx>0$ for all $x \neq 0$, then A is said to be **<u>positive-definite</u>** (i.e., the covariance matrix)
  - if $x^TAx \geq 0$ for all $x \neq 0$, then A is said to be **<u>positive-semidefinite</u>**
- **The <u>inverse</u> of a square matrix A is denoted by A^{-1} and is such that $AA^{-1}= A^{-1}A=I$**
  - The inverse A^{-1} of a matrix A exists **<u>if and only if</u>** A is non-singular
- **The <u>pseudo-inverse</u> matrix $A^{\dagger}$ is typically used whenever A^{-1} does not exist (because A is not square or A is singular):**

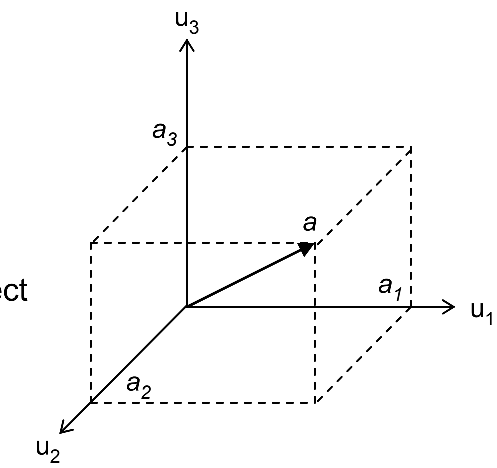
$$A^{\dagger} = \left[A^TA\right]^{-1}A^T \text{ with } A^{\dagger}A = I \qquad \left(\text{assuming } A^TA \text{ is non-singular, note that } AA^{\dagger} \neq I \text{ in general}\right)$$

Vector spaces

- **The n-dimensional space in which all the n-dimensional vectors reside is called a vector space**
- **A set of vectors $\{u_1, u_2, \ldots u_n\}$ is said to form a <u>basis</u> for a vector space if any arbitrary vector x can be represented by a linear combination of the $\{u_i\}$**

$$x = a_1u_1 + a_2u_2 + \cdots a_nu_n$$

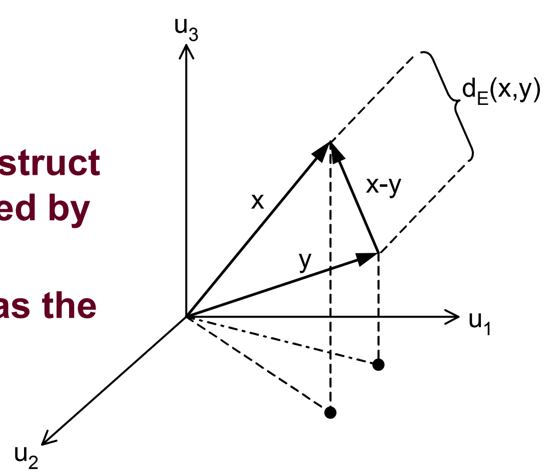
  - The coefficients $\{a_1, a_2, \ldots a_n\}$ are called the <u>components</u> of vector x with respect to the basis $\{u_i\}$
  - In order to form a basis, it is necessary and sufficient that the $\{u_i\}$ vectors be linearly independent



- **A basis $\{u_i\}$ is said to be <u>orthogonal</u> if** $u_i^T u_j \begin{cases} \neq 0 & i = j \\ = 0 & i \neq j \end{cases}$
- **A basis $\{u_i\}$ is said to be <u>orthonormal</u> if** $u_i^T u_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$
  - As an example, the Cartesian coordinate base is an orthonormal base
- **Given n linearly independent vectors $\{x_1, x_2, \ldots x_n\}$, we can construct an orthonormal base $\{\phi_1, \phi_2, \ldots \phi_n\}$ for the vector space spanned by $\{x_i\}$ with the <u>Gram-Schmidt</u> Orthonormalization Procedure**
- **The <u>distance</u> between two points in a vector space is defined as the magnitude of the vector difference between the points**

$$d_E(x,y) = |x - y| = \left[\sum_{k=1}^{d} (x_k - y_k)^2\right]^{1/2}$$

  - This is also called the Euclidean distance



Linear transformations

- **A <u>linear transformation</u> is a mapping from a vector space X^N onto a vector space Y^M, and is represented by a matrix**
  - Given vector $x \in X^N$, the corresponding vector y on Y^M is computed as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1N} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \ddots & \\ a_{M1} & a_{M2} & a_{M3} & & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_N \end{bmatrix}$$

  - Notice that the dimensionality of the two spaces does not need to be the same.
  - For pattern recognition we typically have M<N (project onto a lower-dimensionality space)
- **A linear transformation represented by a square matrix A is said to be <u>orthonormal</u> when $AA^T=A^TA=I$**
  - This implies that $A^T=A^{-1}$
  - An orthonormal transformation has the property of preserving the magnitude of the vectors:

$$|y| = \sqrt{y^T y} = \sqrt{(Ax)^T (Ax)} = \sqrt{x^T A^T A x} = \sqrt{x^T x} = |x|$$

  - An orthonormal matrix can be thought of as a rotation of the reference frame
  - The **row vectors** of an orthonormal transformation form a set of orthonormal basis vectors

$$y_{M\times 1} = \begin{bmatrix} \leftarrow & a_1 & \rightarrow \\ \leftarrow & a_2 & \rightarrow \\ & & \\ \leftarrow & a_N & \rightarrow \end{bmatrix} x_{N\times 1} \quad \text{with} \quad a_i^T a_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Eigenvectors and eigenvalues

- **Given a matrix $A_{N \times N}$, we say that v is an <u>eigenvector</u>* if there exists a scalar λ (the <u>eigenvalue</u>) such that**

$$Av = \lambda v \Leftrightarrow \begin{cases} v \text{ is an eigenvector} \\ \lambda \text{ is the corresponding eigenvalue} \end{cases}$$

- **Computation of the eigenvalues**

$$Av = \lambda v \Rightarrow Av - \lambda v = 0 \Rightarrow (A - \lambda I)v = 0 \Rightarrow \begin{cases} v = 0 & \text{trivial solution} \\ (A - \lambda I) = 0 & \text{non-trivial solution} \end{cases}$$

$$(A - \lambda I) = 0 \Rightarrow |A - \lambda I| = 0 \Rightarrow \underbrace{\lambda^N + a_1\lambda^{N-1} + \cdots a_{N-1}\lambda + a_0 = 0}_{\text{Characteristic Equation}}$$

- **The matrix formed by the column eigenvectors is called the <u>modal matrix</u> M.**
  - Matrix Λ is the <u>canonical form</u> of A: a diagonal matrix with eigenvalues on the main diagonal

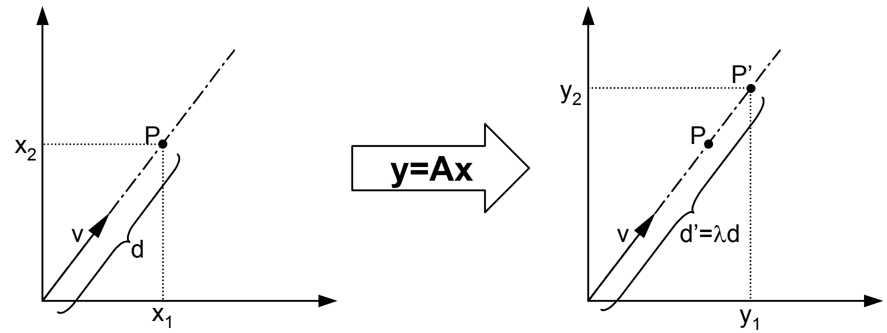
$$M = \begin{bmatrix} \uparrow & \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & v_3 & \cdots & v_N \\ \downarrow & \downarrow & \downarrow & & \downarrow \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix}$$

- **Properties**
  - If A is non-singular
    - All eigenvalues are non-zero
  - If A is real and symmetric
    - All eigenvalues are real
    - The eigenvectors associated with distinct eigenvalues are orthogonal
  - If A is positive definite
    - All eigenvalues are positive

**The "eigen-" in "eigenvector" is normally translated as "characteristic"*

Interpretation of eigenvectors and eigenvalues (1)

- **If we view matrix A as a linear transformation, an eigenvector represents an invariant direction in the vector space**
  - When transformed by A, any point lying on the direction defined by v will remain on that direction, and its magnitude will be multiplied by the corresponding eigenvalue λ



  - For example, the transformation which rotates 3-d vectors about the Z axis has vector [0 0 1] as its only eigenvector and 1 as the corresponding eigenvalue

$$A = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \Leftrightarrow$$