Raied Aljadaany

Pattern Recognition Theory

Recitation 2: Linear Algebra

Electrical & Computer ENGINEERING

The Basics

- Scalar (x)
  - A real number

- Vector ($\mathbf{x}$)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$$

The Basics

- Matrix (**X**)

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & & x_{nm} \end{bmatrix}$$

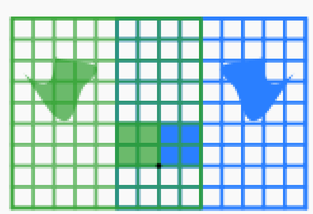
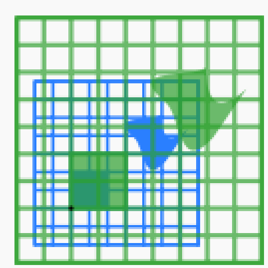
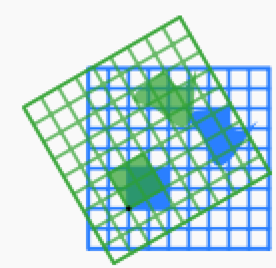
Types of Matrices

- Square & Rectangular matrices
- Upper & Lower Triangular matrices
- Identity matrix
- Symmetric & Skew-symmetric matrices
- Hermitian matrix
- Singular matrix
- Orthogonal matrix

Consider a matrix as…

- … an *operator* …
- … which *linearly transforms* a vector …
- … to a different *vector space*!

For example:

Horizontal flip	Scaling by a factor of 3/2	Rotation by π/6R = 30°
$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3/2 & 0 \\ 0 & 3/2 \end{bmatrix}$	$\begin{bmatrix} \cos(\pi/6^R) & -\sin(\pi/6^R) \\ \sin(\pi/6^R) & \cos(\pi/6^R) \end{bmatrix}$
		

* Examples from Wikipedia

Vector Spaces & Subspaces

For any m x n matrix **A** with rank r:

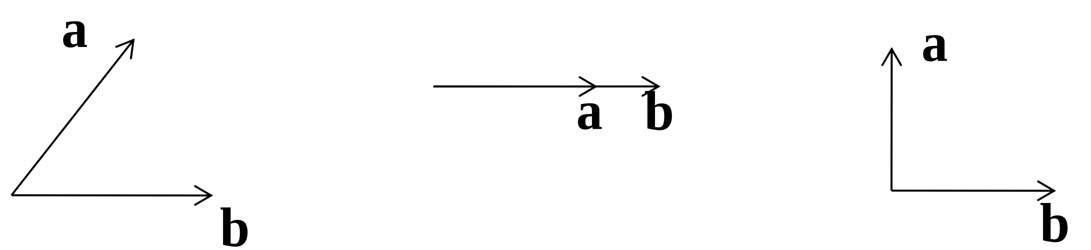
- Column Space
  - Contains all linear combinations of the columns of the matrix
  - Has dimension r
  - **Ax** = **b** can be solved iff **b** is in the column space of **A.**
- Null Space
  - Contains all vectors **x** such that **Ax** = 0
  - Has dimension (n-r)
- Row Space / Left Null Space (equivalent to Column & Null Space on $\mathbf{A}^T$)

Inner Products

- Inner product of **x** and **y**: Sum of element-wise products

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

- **x** and **y** must be equal in length
- Result is a **scalar**
  - Test of similarity of two vectors
  - Don't forget to normalize vectors before comparing!



Outer Products

- Outer product of **x** and **y**:

$$\mathbf{x} \otimes \mathbf{y} = \mathbf{x}\mathbf{y}^{\mathbf{T}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & x_1 y_3 \\ x_2 y_1 & x_2 y_2 & x_2 y_3 \\ x_3 y_1 & x_3 y_2 & x_3 y_3 \\ x_4 y_1 & x_4 y_2 & x_4 y_3 \end{bmatrix}$$

- **x** and **y** can be of different lengths
- Result is a matrix

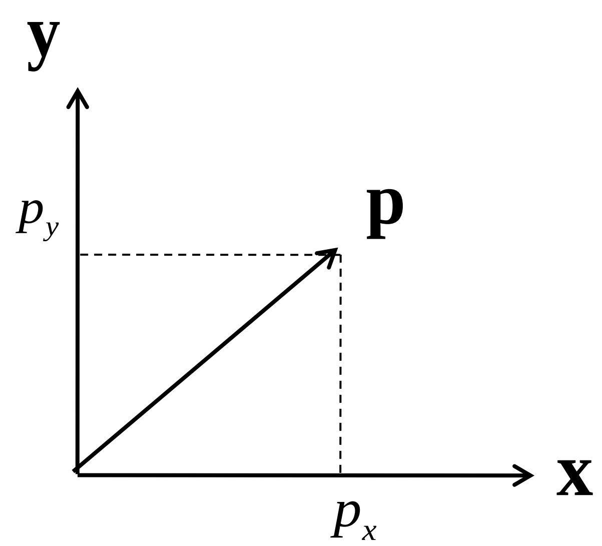
Linear independence

- Non-zero vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ are linearly independent only if

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 \neq 0$$

  for any non-zero set of constants a_n
- We say a family of vectors is a linearly independent family if none of them can be written as a linear combination of finitely many other vectors in the family.

Linear independence & Orthogonality



Inner product of orthogonal vectors is 0

$$\mathbf{x}^T \mathbf{y} = \mathbf{0}$$

Projections

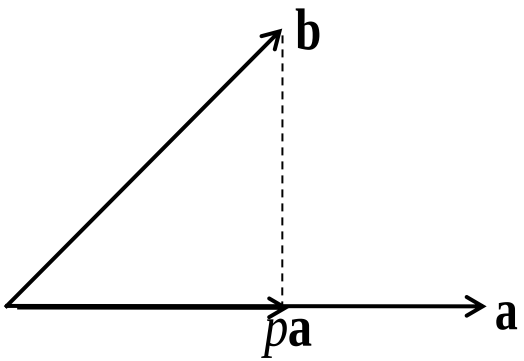
- We want to find the value of *p* which minimizes the error ||**b** - *p***a**||, i.e.

$$\hat{p} = \arg\min_p \|\mathbf{b} - p\mathbf{a}\|$$
$$= \arg\min_p (\mathbf{b} - p\mathbf{a})^T(\mathbf{b} - p\mathbf{a})$$
$$= \arg\min_p (\mathbf{b}^T\mathbf{b} + p^2\mathbf{a}^T\mathbf{a} - 2p\mathbf{a}^T\mathbf{b})$$

Taking derivative and setting it to 0 :

$$0 = 2\hat{p}\mathbf{a}^T\mathbf{a} - 2\mathbf{a}^T\mathbf{b}$$

$$\hat{p} = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$$



- Hence, if **a** is unit norm, the projection coefficient is equivalent to the inner product of **a** and **b**

Matrix Operations

1. Matrix <u>Transpose</u>, Conjugate Transpose
   - $(\mathbf{A}^T)^T = \mathbf{A}$
   - $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
   - $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
   - $(r\mathbf{A})^T = r\mathbf{A}^T$

2. Matrix <u>Determinant</u>
   - Only exists for square matrices
   - $\det(a\mathbf{X}) = a^n\det(\mathbf{X})$
   - $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
   - $\det(\mathbf{A}^T) = \det(\mathbf{A})$
   - $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$

Matrix Operations

3. Matrix arithmetic:
   - Addition & Subtraction: Element-wise operations
   - Matrix Multiplication:
     - $\mathbf{AB} \neq \mathbf{BA}$
   - Vector Multiplication
   - Scalar Multiplication

4. Matrix Inverse:
   - $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
   - $(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}$

5. Matrix Rank

Matrix Operations

6. (Moore-Penrose) Pseudo-Inverse :
   - The pseudoinverse A^+ of a matrix A is a generalization of the inverse matrix. The most widely known is the Moore-Penrose pseudoinverse.
   - The pseudoinverse is defined and unique for all matrices whose entries are real or complex numbers.
   - If the matrix A has dimensions m x n, and is full rank, then use the left inverse if m>n, and the right inverse if m<n.
   - Left inverse: $A_{\mathrm{left}}^{-1} = \left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$, i.e. $A_{\mathrm{left}}^{-1}A = I_n$
   - Right inverse: $A_{\mathrm{right}}^{-1} = A^{\mathrm{T}}\left(AA^{\mathrm{T}}\right)^{-1}$, i.e. $AA_{\mathrm{right}}^{-1} = I_m$

System of Linear Equations

- Suppose we have a set of linear equations such as the following example:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3
\end{aligned}
$$

In matrix form, we can write this as:

$$
\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}
=
\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}
$$

$$
\mathbf{A}\,\mathbf{x} = \mathbf{b}
$$

System of Linear Equations Ax = b

	Under-determined	Well defined	Over-determined
Equations vs. Unknowns	Less linearly independent equations than unknowns	As many linearly independent equations as unknowns	More linearly independent equations than unknowns
A	**A** is "fat"	**A** is square	**A** is "tall"
# of solutions	Usually infinitely many solutions	Usually one solution	Usually no solution
Typical solution	Minimum Norm solutions $\mathbf{x} = \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$	<u>Exact</u> solution: $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$	<u>*Least squared error*</u> solution: $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$

Eigen decomposition

- Any matrix **A** represents a transformation operation on a vector

$$\mathbf{A}\mathbf{x} = \mathbf{x}'$$

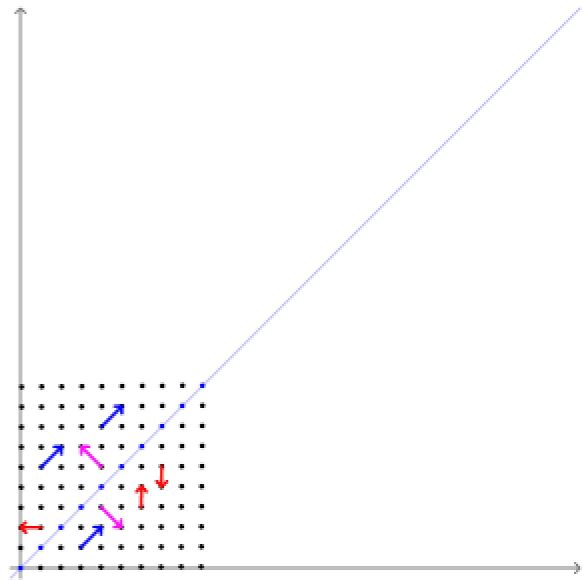
- For certain vectors, the transformation is merely a scale change

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

Eigen decomposition

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

- $\mathbf{x}$ is the eigen vector of transformation $\mathbf{A}$
- λ is the corresponding eigen value



Determining the eigen decomposition

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

$$\left(\mathbf{A} - \lambda\mathbf{I}\right)\mathbf{x} = 0$$

For a non-trivial solution

$$\left|\mathbf{A} - \lambda\mathbf{I}\right| = 0$$

This equation is called the characteristic equation.
Solve for the eigen values λ and hence obtain '$\mathbf{x}$'s

Some properties of eigen decomposition

- Rank of **A** = Number of non-zero eigen values of **A** = Number of linearly independent eigen vectors

- Determinant of $\mathbf{A}$, $|\mathbf{A}| = \prod_{i=1}^{i=n} \lambda_i$

- Trace of $\mathbf{A} = \sum_{i=1}^{i=n} \lambda_i$

Some properties of eigen decomposition

- If **A** is <u>symmetric</u>, then the eigen vectors are orthogonal
- If λ is an eigen value of **A**,
  - then $\frac{1}{\lambda}$ is an eigen value of $\mathbf{A}^{-1}$
  - $\lambda - k$ is an eigen value of $\mathbf{A} - k\mathbf{I}$
  - λ^m is an eigen value of $\mathbf{A}^m$

Eigen decomposition

- A matrix is
  - positive definite if all its eigen values

$$\lambda_i > 0$$

  - positive semi-definite if

$$\lambda_i \geq 0$$

  - negative definite if

$$\lambda_i < 0$$

  - negative semi-definite if

$$\lambda_i \leq 0$$

Calculus

$$\nabla f(\mathbf{x}) = gradient(f(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

$$\boxed{\begin{array}{l} \frac{\partial \mathbf{x}^{\mathrm{T}}\mathbf{c}}{\partial \mathbf{x}} = \frac{\partial \mathbf{c}^{\mathrm{T}}\mathbf{x}}{\partial \mathbf{x}} = c \\ \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A} \end{array}}$$

Product Rule: $\frac{\partial (g(\mathbf{x}))^{\mathrm{T}} h(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial (h(\mathbf{x}))^{\mathrm{T}}}{\partial \mathbf{x}} g(\mathbf{x}) + \frac{\partial (g(\mathbf{x}))^{\mathrm{T}}}{\partial \mathbf{x}} h(\mathbf{x})$

e.g.: $\frac{\partial \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A}^{\mathrm{T}} + \mathbf{A})\mathbf{x}$ In this case $(g(\mathbf{x}) = \mathbf{x}, h(\mathbf{x}) = \mathbf{A}\mathbf{x})$

Chain Rule: $\frac{\partial g(h(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial g}{\partial h} \cdot \frac{\partial h}{\partial \mathbf{x}}$

Highly Recommended

- "Linear Algebra and its applications" – 4th Ed., Gilbert Strang
- The Matrix Cookbook
  - http://matrixcookbook.com/
- Gilbert Strang MIT video lectures
  - http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/video-lectures/