Linear Algebra Review

Basics only…

Taken mostly from G. Bebis

1

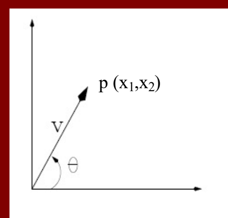
Review topics

- Vectors
- Matrices
- Eigenvalues/Eigenvectors
- Linear systems of equations

2

2D vectors

- Represented geometrically as directed line segments.
- Two attributes: *magnitude* and *direction*



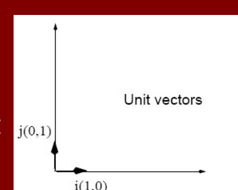
representation: $v = (x_1, x_2)$

magnitude: $||v|| = \sqrt{x_1^2 + x_2^2}$

direction: $\theta = \tan\ (x_2/x_1)$

3

Unit vectors



- Any vector with magnitude equal to one.
- E.g., unit vectors in the *x* and *y* directions:

  $i = (1, 0),\ j = (0, 1)$

- Vector normalization $\bar{v} = \frac{v}{||v||} = (\frac{x_1}{||v||}, \frac{x_2}{||v||})$

Example: $v = (2, 5),\ ||v|| = \sqrt{2^2 + 5^2} = \sqrt{29}$

$\bar{v} = (\frac{2}{\sqrt{29}}, \frac{5}{\sqrt{29}})$ unit vector

4

Vector arithmetic

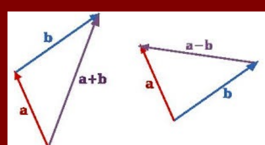
- Addition:

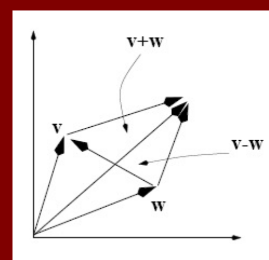
  $v + w = (v_1, v_2) + (w_1, w_2) = (v_1 + w_1, v_2 + w_2)$
- Subtraction:

  $v - w = (v_1, v_2) - (w_1, w_2) = (v_1 - w_1, v_2 - w_2)$
- Scalar multiplication:

  $av = (av_1, av_2)$

  (a – scalar)





5

n-dimensional vector notation

- An n-dimensional vector v and its transpose v^T can be denoted as follows:

$$v = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \qquad v^T = [\, x_1 \; x_2 \; \cdots \; x_n \,]$$

6

Inner (or dot) product

- Given two vectors $v = (x_1, x_2, \ldots, x_n)$ and $w = (y_1, y_2, \ldots, y_n)$, their dot product defined as follows:

$$v.w = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$ scalar!

or

$$v.w = [\, x_1 \; x_2 \; \cdots \; x_n \,] \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = v^T w$$

7

Magnitude using dot product notation

- The magnitude (or *Euclidean norm)* of an n-dimensional vector is defined as:

$$||v|| = \sqrt{v^T v} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$v.v = [x_1 \; x_2 \; \cdots \; x_n] \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + \ldots + x_n^2 = ||v||^2$$

8

Outer product

$$wv^T = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} [\, x_1 \; x_2 \cdots x_n \,] = \begin{bmatrix} y_1x_1 & y_1x_2 & \dots & y_1x_n \\ y_2x_1 & y_2x_2 & \dots & y_2x_n \\ \dots & \dots & \dots & \dots \\ y_nx_1 & y_nx_2 & \dots & y_nx_n \end{bmatrix}$$

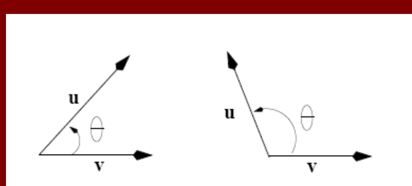
Warning: the inner product is a **scalar**,
the outer product is a **matrix** !

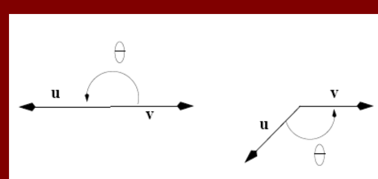
9

Geometric definition of dot product

$$u.\, v = \|u\| \; \|v\| \; cos(\theta),\; 0 \le \theta \le \pi$$

θ corresponds to the smaller angle between u and v

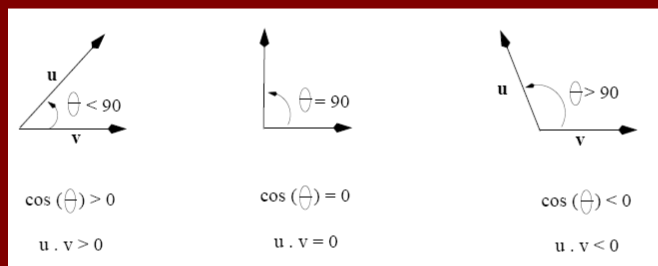




The dot product is fundamentally a projection (dot product of a vector with a unit vector is the projection of that vector in the direction given by the unit vector).

10

Geometric definition of dot product (cont' d)



- The sign of u.v depends on cos(θ)
- $v.\ (u + aw) = v.\ u + a(v.\ w)$

11

Orthogonal/Orthonormal vectors

- A set of vectors $x_1, x_2, \ldots, x_n$ is *orthogonal (perpendicular)* if

$$x_i^T x_j = \begin{cases} k & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

- A set of vectors $x_1, x_2, \ldots, x_n$ is *orthonormal (orthogonal and unit vectors)* if

$$x_i^T x_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

12

Linear combinations of vectors

- A vector v is a linear combination of the vectors $v_1, \ldots, v_k$:

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

where $c_1, \ldots, c_k$ are scalars

- e.g., vectors in R^3 are linear combinations of the unit vectors $i = (1, 0, 0)$, $j = (0, 1, 0)$, and $k = (0, 0, 1)$

$$v = (a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1)$$

13

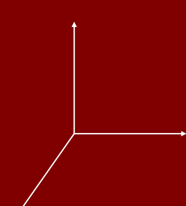
Space spanning

- A set of vectors $S = (v_1, v_2, \ldots, v_k)$ *span* some space W if every vector in W can be written as a linear combination of the vectors in S

$$\mathrm{w} = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

- e.g., the vectors i, j, and k span R^3

$$v = (a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1)$$



14

Linear independence

- A set of vectors $v_1, \ldots, v_k$ is *linearly independent* if

$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = 0$ implies $c_1 = c_2 = \cdots = c_k = 0$

Example:

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Let $c_1 v_1 + c_2 v_2 = 0$, then $\begin{bmatrix} -c_1 + c_2 \\ c_1 + c_2 \\ -c_1 + (-c_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

This can only be true if $c_1 = c_2 = c_3 = 0$

15

Linear dependence

- A set of vectors $v_1, \ldots, v_k$ is *linearly dependent* if at least one of them is a linear combination of the others.

$$v_j = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

(i.e., v_j does not appear at the right side)

16

Geometric interpretation of linear dependence/independence

- In R^2, two vectors are linearly independent if they do not lie on the same line.
- In R^3, three vectors are linearly independent if not all of them lie on the same plane.

17

Vector basis

- A set of vectors $S=(v_1, ..., v_k)$ is said to be a *basis* for a vector space W if
  - **(1)** v_j are linearly independent
  - **(2)** S spans W
- Standard bases:

R^2	R^3	R^n
$i = (1, 0),\ j = (0, 1)$	$i = (1, 0, 0),\ j = (0, 1, 0),\ k = (0, 0, 1)$	$(1,0,...,0)\ (0,1,...,0),\ ....,\ (0,0,...,1)$

18

Space dimension

- There exist multiple sets of basis vectors in a given space.
- Each base in *W* has the <u>same</u> number of vectors.
- The <u>dimension</u> of a vector space *W* is determined by the size of any one of its sets of basis vectors.

19

Gram-Schmidt orthogonalization

- Any set of basis vectors ($x_1, x_2, \ldots, x_n$) can be transformed to an orthogonal basis ($o_1, o_2, \ldots, o_n$) using the *Gram-Schmidt* orthogonalization procedure.

  (1) Take $o_1 = x_1$

  (2) Take the second vector x_2 and subtract from it the part that lies along the direction o_1

$$o_2 = x_2 - a o_1, \text{ where } a = \frac{x_2 . o_1}{o_1 . o_1}$$

  (3) Continue the process to obtain all the vectors:

$$o_k = x_k - \sum_{i=1}^{k-1} \frac{x_k . o_i}{o_i . o_i} o_i$$

20

Vector expansion

- Suppose $v_1, v_2, \ldots, v_n$ represents a base in W, then any $v \in W$ has a **unique** vector expansion in this base:

$$v = \sum_{i=1}^{n} x_i v_i = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n$$

- The vector expansion provides a meaning for writing a vector as a "column of numbers".

$$v = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

Note: to interpret v, we need to know what basis was used for the expansion !!

$$v = (a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1)$$

21

How to compute the coefficients of vector expansion?

(1) Assuming the basis vectors are orthogonal, to compute x_i, take the inner product of v_i and v

$$v_i . v = v_i . \sum_{j=1}^{n} x_j v_j = \sum_{j=1}^{n} x_j v_j . v_i = x_i v_i . v_i$$

(2) The coefficients of the expansion can be computed as follows:

$$x_i = \frac{v . v_i}{v_i . v_i}$$

22

Review on Matrices

- Matrix addition/subtraction
  - Matrices can be added or subtracted as long as they are of the same size.
- Matrix multiplication

$$\underset{m \times n}{\begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & . & a_{mn} \end{bmatrix}} \underset{q \times p}{\begin{bmatrix} b_{11} & b_{12} & . & b_{1p} \\ b_{21} & b_{22} & . & b_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ b_{q1} & b_{q2} & . & b_{qp} \end{bmatrix}} = \underset{m \times p}{\begin{bmatrix} c_{11} & c_{12} & . & c_{1p} \\ c_{21} & c_{22} & . & c_{2p} \\ \cdots & \cdots & c_{ij} & \cdots \\ c_{m1} & c_{m2} & . & c_{mp} \end{bmatrix}}$$

Condition: n = q

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}$$

23

Properties of matrix multiplication

$$A(B+C) = AB + AC \text{ (distributive law)}$$

$$AB \neq BA$$

unit matrix

$$AI = IA = A, \text{ where } I = \begin{bmatrix} 1 & 0 & . & 0 \\ 0 & 1 & . & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & . & 1 \end{bmatrix}$$

24

Matrix Transpose

$$A = \begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & . & a_{mn} \end{bmatrix}, \quad A^T = \begin{bmatrix} a_{11} & a_{21} & . & a_{m1} \\ a_{12} & a_{22} & . & a_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & . & a_{mn} \end{bmatrix}$$

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

25

Symmetric Matrices

$$A = A^T \quad (a_{ij} = a_{ji})$$

Example:

$$\begin{bmatrix} 4 & 5 & -3 \\ 5 & 7 & 2 \\ -3 & 2 & 10 \end{bmatrix}$$

26

Determinants

2 x 2

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

3 x 3

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

n x n

$$det(A) = \sum_{j=1}^{m} (-1)^{j+k} a_{jk} det(A_{jk}), \text{ for any } k\text{: } 1 \le k \le m$$

27

Determinants (cont' d)

$$det(AB) = det(A)det(B)$$

$$det(A+B) \neq det(A) + det(B)$$

diagonal matrix:

$$\text{If } A = \begin{bmatrix} a_{11} & 0 & . & 0 \\ 0 & a_{22} & . & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & . & a_{nn} \end{bmatrix}, \text{ then } det(A) = \prod_{i=1}^{n} a_{ii}$$

28

Matrix Inverse

- The inverse A^{-1} of a matrix A has the property:

$$AA^{-1}=A^{-1}A=I$$

- A^{-1} exists only if $det(A) \neq 0$
- **Singular matrix:** A^{-1} does not exist
- **Ill-conditioned matrix:** A is close to being singular
- Properties of the inverse:

$$det(A^{-1}) = \frac{1}{det(A)}$$

$$(AB)^{-1}=B^{-1}A^{-1}$$

$$(A^T)^{-1} = (A^{-1})^T$$

29

Pseudo-inverse

- The pseudo-inverse A^+ of a matrix A (could be non-square, i.e., m x n) is given by:

$$A^+ = (A^T A)^{-1} A^T$$

- It can be shown that:

$$A^+ A = I \quad \text{(provided that } (A^T A)^{-1} \text{ exists)}$$

*(If the **rank is < matrix dimension**, or **matrix is not square**)*

30

Matrix trace

$$tr(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

properties:

$$tr(A^T) = tr(A)$$

$$tr(A \pm B) = tr(A) \pm tr(B)$$

$$tr(AB) = tr(BA)$$

$$(\text{in general, } tr(AB) \neq tr(A)tr(B))$$

31

Rank of matrix

- Equal to the dimension of the largest square submatrix of A that has a non-zero determinant.

Example: $\begin{bmatrix} 4 & 5 & 2 & 14 \\ 3 & 9 & 6 & 21 \\ 8 & 10 & 7 & 28 \\ 1 & 2 & 9 & 5 \end{bmatrix}$ has rank 3

$$det(A) = 0, \text{ but } det(\begin{bmatrix} 4 & 5 & 2 \\ 3 & 9 & 6 \\ 8 & 10 & 7 \end{bmatrix}) = 63 \neq 0$$

32

Rank of matrix (cont' d)

- **Alternative definition:** the maximum number of **linearly independent columns (or rows)** of A.

$$1\begin{bmatrix}4\\3\\8\\1\end{bmatrix}+2\begin{bmatrix}5\\9\\10\\2\end{bmatrix}+0\begin{bmatrix}2\\6\\7\\9\end{bmatrix}-1\begin{bmatrix}14\\21\\28\\5\end{bmatrix}=0$$

Rank is not 4 !

Using rank, we will be able to define sets of basis vectors ….

33

Rank and singular matrices

If A is nxn, $rank(A) = n$ iff A is nonsingular (i.e., invertible).

If A is nxn, $rank(A) = n$ iff $det(A) \neq 0$ (**full rank**).

If A is nxn, $rank(A) < n$ iff A is singular

(Singular matrix: iff its determinant =0)

34

Orthogonal matrices

- Notation:

$$A = \begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & . & a_{mn} \end{bmatrix},$$

$$\begin{aligned} u_1^T &= [a_{11}\ a_{12}\ \cdots\ a_{1n}] \\ u_2^T &= [a_{21}\ a_{22}\ \cdots\ a_{2n}] \\ &\cdots \\ u_m^T &= [a_{m1}\ a_{m2}\ \cdots\ a_{mn}] \end{aligned}$$

$$A = \begin{bmatrix} u_1^T \\ u_2^T \\ \cdots \\ u_m^T \end{bmatrix}$$

- A is orthogonal if:

$$u_j . u_k = 0, \text{ for every } j \neq k \ (u_j \text{ is perpendicular to } u_k)$$

Example: $\begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix}$

35

Orthonormal matrices

- A is orthonormal if:

(1) $u_k . u_k = 1$ or $||u_k|| = 1$, for every k

(2) $u_j . u_k = 0$, for every $j \neq k$ (u_j is perpendicular to u_k)

- If A is orthonormal, then it easy to find its inverse:

$$AA^T = A^T A = I \quad (\text{i.e., } A^{-1} = A^T)$$

also: $||Av|| = ||v||$ (does not change the magnitude of v)

36

Eigenvalues and Eigenvectors

- The vector **v** is an eigenvector of matrix A and $\boldsymbol{\lambda}$ is an eigenvalue of A if:

$$Av = \lambda v$$ (assume non-zero v)

- Interpretation: the linear transformation implied by A cannot change the direction of the eigenvectors v, only their magnitude.

37

Characteristic polynomial

- To find the eigenvalues λ of a matrix A, find the roots of the *characteristic polynomial*:

$$det(A - \lambda I) = 0$$

Example: $A = \begin{bmatrix} 5 & -2 \\ 6 & -2 \end{bmatrix}$

$$det(\begin{bmatrix} 5-\lambda & -2 \\ 6 & -2-\lambda \end{bmatrix}) = 0 \text{ or } \lambda^2 - 3\lambda + 2 = 0 \text{ or } \lambda_1 = 1,\ \lambda_2 = 2$$

$$v_1 = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix},\ v_2 = \begin{bmatrix} 2/3 \\ 1 \end{bmatrix}$$

38

Properties

- Eigenvalues and eigenvectors are only defined for square matrices (i.e., $m = n$)
- Eigenvectors are not unique (e.g., if v is an eigenvector, so is kv)

$$Av = \lambda v$$

- Suppose $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A, then:

(1) $\sum_i \lambda_i = tr(A)$

(2) $\prod_i \lambda_i = det(A)$

(3) if $\lambda = 0$ is an eigenvalue, then the matrix is not invertible

39

Properties (cont' d)

(4) A and A^2 have the same eigenvectors

(5) if λ is an eigenvalue of A, then λ^2 is an eigenvalue of A^2

(6) a matrix A with positive eigenvalues is called *positive definite*

$x^T A x > 0$ for every $x \neq 0$

40

Diagonalization

- Given A, find P such that $P^{-1}AP$ is diagonal (i.e., we say that P diagonalizes A)
- Take $P = [v_1\ v_2\ .\ .\ .\ v_n]$, where $\lambda_1, \lambda_2, ..., \lambda_n$ are the eigenvalues of A and $v_1, v_2, .\ .\ .\ v_n$ are the eigenvectors of A:

$$Av = \lambda v \longrightarrow AP = P\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & \lambda_n \end{bmatrix} \quad \text{or} \quad P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & \lambda_n \end{bmatrix}$$

41

Diagonalization (cont' d)

Example: $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

$$\lambda_1 = 0,\ \lambda_2 = 2,\ v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix},\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$P = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},\ P^{-1} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

$$P^{-1}AP = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

42

Are all n x n matrices diagonalizable?

- Only if P^{-1} exists! - *A* must have *n* linearly independent eigenvectors;

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ . & . & . \\ . & . & . \\ . & . & . \\ 0 & 0 & \lambda_n \end{bmatrix}$$

- If *A* has *n* distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$, then the corresponding eigevectors $v_1, v_2, . . . v_n$ are:
  - **(1)** linearly independent and
  - **(2)** span R^n (i.e., form a basis)

43

Diagonalization → Decomposition

- Let us assume that *A* is diagonalizable, then:

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ . & . & . \\ . & . & . \\ . & . & . \\ 0 & 0 & \lambda_n \end{bmatrix} \longrightarrow A = P \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ . & . & . \\ . & . & . \\ . & . & . \\ 0 & 0 & \lambda_n \end{bmatrix} P^{-1}$$

44

Decomposition: symmetric matrices

The eigenvalues of symmetric matrices are all real and the eigenvectors corresponding to distinct eigenvalues are orthogonal (or orthonomal by normalizing them).

$$A = P \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & \lambda_n \end{bmatrix} P^{-1} \longrightarrow$$

$P^{-1}=P^T$ symmetric matrices:

$$A=PDP^T=\sum_{i=1}^{n} \lambda_i v_i v_i^T$$

45

Systems of linear equations

- An arbitrary system of m linear equations in n unknowns can be written as:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

or $Ax = b$ where:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

$$\begin{aligned} 2x_1 + 3x_2 &= 8 \\ 4x_1 + 6x_2 &= 9 \end{aligned} \longrightarrow \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 9 \end{bmatrix}$$

46

Over/Under determined Systems

When m>n the system is called *over-determined.*

When m<n the system is called *under-determined.*

$$
\begin{gathered}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\
\ldots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m
\end{gathered}
$$

47

Solving Ax=b (case m=n, square matrix)

(1) If A is invertible (includes, being square), $Ax = b$ has exactly *one solution*:

$$x = A^{-1}b$$

- The following statements are equivalent:
  - (a) A is invertible
  - (b) $det(A) \neq 0$
  - (c) $Ax = 0$ has the trivial solution only
  - (d) b is in the column space of A *(linear combination of columns)*

$$
\begin{bmatrix} a_{11} \\ a_{21} \\ \ldots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ \ldots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \ldots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ b_n \end{bmatrix}
$$

48

Solving Ax=b (case m=n)

(e) $rank(A|b) = rank(A)$ and $rank(A) = n$ (i.e., from (d))
(f) The column/row vectors of A are linearly independent.
(g) The column/row vectors of A span R^n

(2) The system has *no solution* if

$$rank(A|b) > rank(A)$$

(3) The system has *infinitely many solutions* if

$$rank(A|b) = rank(A) \text{ and } rank(A) < n$$

49

Classical (non)Linear Regression (i.e., Classification)

n = number of data, m = number of features, attributes, inputs
In a linear model entries x_{ij} of $\boldsymbol{X}$ are features x_{ij}, and in a nonlinear case entries are some functions of **x**, $x_{ij} = f(x_{ij})$

$$X_{nm}\vec{w}_{m1} = \vec{y}_{n1} \qquad /*_{left}\ X^T_{mn}$$

$$X^T_{mn}X_{nm}\vec{w}_{m1} = X^T_{mn}\vec{y}_{n1}$$

NORMAL SYSTEM

$$\left(X^T_{mn}X_{nm}\right)\vec{w}_{m1} = X^T_{mn}\vec{y}_{n1}$$

$$\left(X^T_{mn}X_{nm}\right)^{-1}\left(X^T_{mn}X_{nm}\right)\vec{w}_{m1} = \left(X^T_{mn}X_{nm}\right)^{-1}X^T_{mn}\vec{y}_{n1}$$

$$\vec{w}_{m1} = \boxed{\left(X^T_{mn}X_{nm}\right)^{-1}X^T_{mn}}\,\vec{y}_{n1}$$

$$\hat{\vec{y}}_{test,1} = X_{test,m}\vec{w}_{m1}$$

Pseudoinverse

Usually, $n > m$ -> overdetermined system

But not always!!!

Check the differences in a meaning of the solution **w**

50

One should be very cautious while using pseudoinverse in solving over- or underdetermined systems

- While for overdetermined systems a solution has a clear meaning a, so called, minimal sum of square errors
- The meaning of the solution for an underdetermined system is geometrically not obvious
- It is only clear that (out of an infinite number of solutions) pseudoinverse will produce an vector **w** with a minimal norm.

51