UVA CS 4501 - 001 / 6501 – 007

Introduction to Machine Learning and Data Mining

Lecture 2:

Yanjun Qi / Jane

University of Virginia
Department of
Computer Science

Yanjun Qi / UVA CS 4501-01-6501-07

Last Lecture Recap

- ❑ Course Logistics
- ❑ My background
- ❑ Basics of machine learning & Application
- ❑ Application and History of MLDM

BASICS OF MACHINE LEARNING

- "The goal of machine learning is to build computer systems that can learn and adapt from their experience." – Tom Dietterich
- "Experience" in the form of available data examples (also called as instances, samples)
- Available examples are described with properties (data points in feature space X)

e.g. SUPERVISED LEARNING

- Find function to map input space X to output space Y $f : X \longrightarrow Y$
- So that the difference between *y* and *f(x)* of each example ***x*** is small.

e.g.

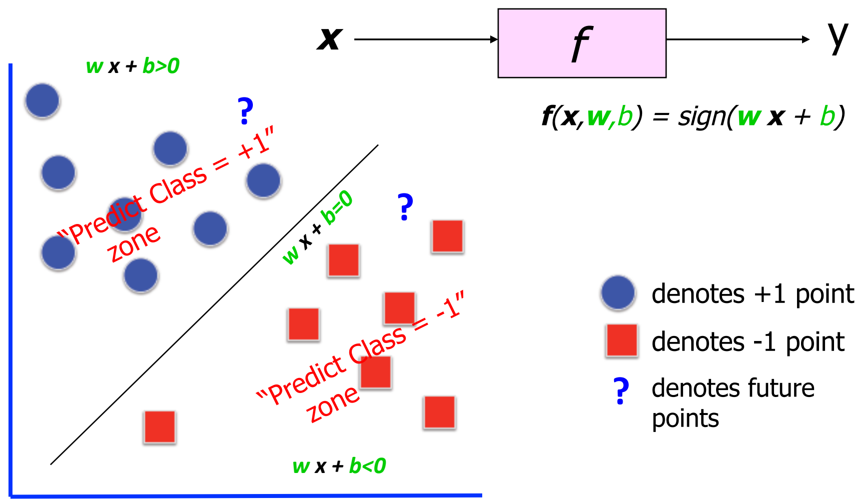
x I believe that this book is not at all helpful since it does not explain thoroughly the material . it just provides the reader with tables and calculations that sometimes are not easily understood …

Input X : e.g. a piece of English text

→ **y** -1

Output Y: {1 / Yes , -1 / No }
e.g. Is this a positive product review

e.g. SUPERVISED Linear Binary Classifier



Courtesy slide from Prof. Andrew Moore's tutorial

Basic Concepts

- Training (i.e. learning parameters $\boxed{\mathbf{w},b}$)
  - Training set includes
    - available examples $\mathbf{x}_1,\ldots,\mathbf{x}_L$
    - available corresponding labels $y_1,\ldots,y_L$
  - Find ($\mathbf{w},b$) by minimizing loss (i.e. difference between y and $f(\mathbf{x})$ *on available examples in training set*)

$$(\mathbf{W}, b) = \underset{\mathbf{W},\, b}{\operatorname{argmin}} \sum_{i=1}^{L} \ell(f(x_i), y_i)$$

Basic Concepts

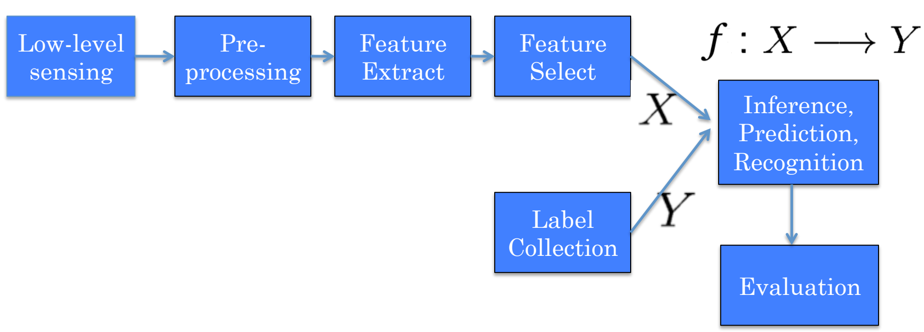
- Testing (i.e. evaluating performance on "future" points)
  - Difference between true $y_?$ and the predicted $f(\boldsymbol{x}_?)$ on a set of testing examples (i.e. *testing set*)
  - Key: example $\boldsymbol{x}_?$ not in the training set

KEY

- Generalisation: learn function / hypothesis from past data in order to "explain", "predict", "model" or "control" new data examples

TYPICAL MACHINE LEARNING SYSTEM

$f : X \longrightarrow Y$

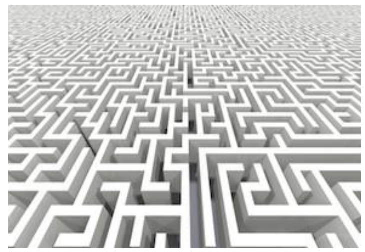


Yanjun Qi / UVA CS 4501-01-6501-07

BIG DATA CHALLENGES FOR MACHINE LEARNING

LARGE-SCALE

Highly Complex



Most of this course

The situations / variations of both X (feature, representation) and Y (labels) are complex !

- ✓Variation of X
- ✓Variation of Y
- ✓ Variation of f

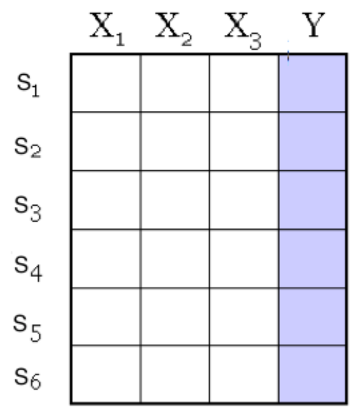
8/28/14 9

Yanjun Qi / UVA CS 4501-01-6501-07

Today:

- ❑ **Data Representation**
- ❑ **Linear Algebra Review**

8/28/14 10

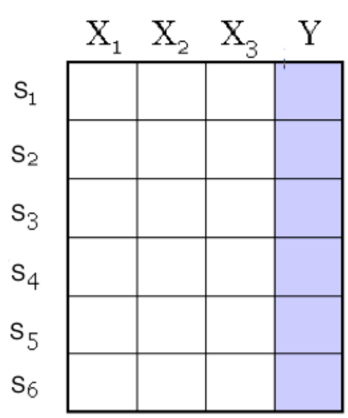


A Dataset

$$f : X \longrightarrow Y$$

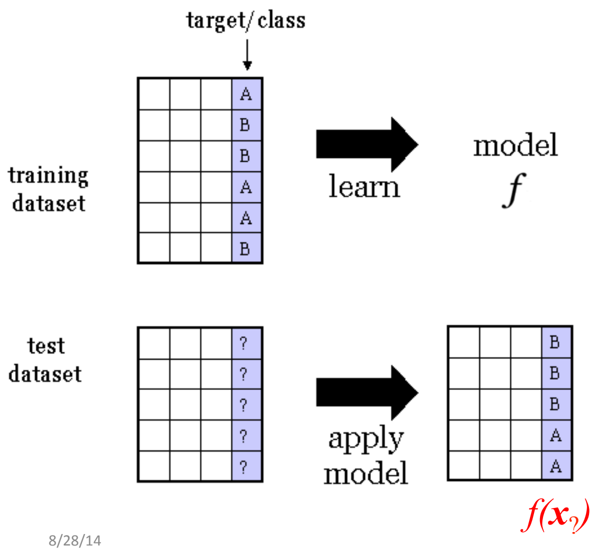
- **Data**/*points/instances/examples/samples/records*: [rows]
- **Features**/*attributes/dimensions/independent variables/covariates/predictors/regressors*: [columns, except the last]
- **Target**/*outcome/response/label/dependent variable*: special column to be predicted [last column]

Main Types of Columns



- *Continuous*: a real number, for example, age or height
- *Discrete*: a symbol, like “Good” or “Bad”

e.g. SUPERVISED LEARNING



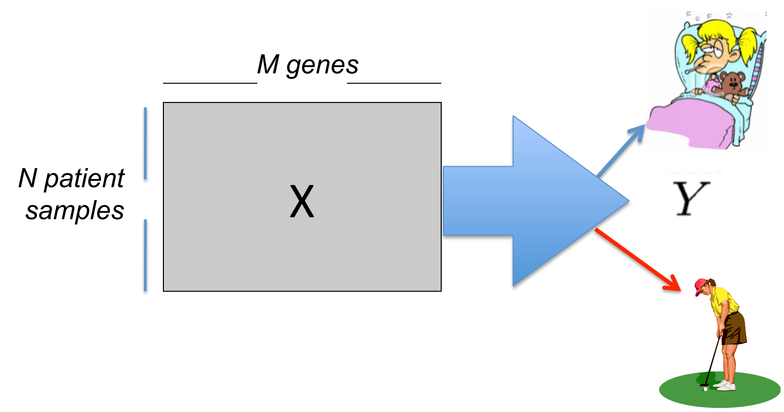
$$f : \boxed{X} \longrightarrow \boxed{Y}$$

- e.g. Target Y is a discrete target variable

Example: Cancer Classification

- Application: automatic disease detection
- Importance: this is modern/future medical diagnosis.
- Prediction goal: Based on past patients, predict whether a new patient have the disease or not
- *Data*: Past patients with and without the disease
- *Target*: Cancer or no-cancer
- *Features*: Concentrations of various genes in your blood --- real values

Example: Cancer Classification



Today:

- ❑ **Data Representation**
- ❑ **Linear Algebra Review**

DEFINITIONS - SCALAR

- a scalar is a number
  - (denoted with regular type: 1 or 22)

DEFINITIONS - VECTOR

- Vector: a single row or column of numbers
  - denoted with **bold small letters**
  - row vector

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

  - column vector

$$\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

DEFINITIONS - VECTOR

- Vector in R^n is an ordered set of n real numbers.
  - e.g. **v** = (1,6,3,4) is in R^4
  - A column vector: $\begin{pmatrix} 1 \\ 6 \\ 3 \\ 4 \end{pmatrix}$
  - A row vector: $\begin{pmatrix} 1 & 6 & 3 & 4 \end{pmatrix}$

DEFINITIONS - MATRIX

◆ A matrix is an array of numbers

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

◆ Denoted with a **bold Capital letter**

◆ All matrices have an order (or dimension): that is, the number of rows × the number of columns. So, **A** is 2 by 3 or (2 × 3).

◆ A square matrix is a matrix that has the same number of rows and columns (n × n)

DEFINITIONS - MATRIX

- m-by-n matrix in R^{mxn} with m rows and n columns, each entry filled with a (typically) real number:
- e.g. 3*3 matrix $\begin{pmatrix} 1 & 2 & 8 \\ 4 & 78 & 6 \\ 9 & 3 & 2 \end{pmatrix}$

Special matrices

$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$ diagonal

$\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$ upper-triangular

$\begin{pmatrix} a & b & 0 & 0 \\ c & d & e & 0 \\ 0 & f & g & h \\ 0 & 0 & i & j \end{pmatrix}$ tri-diagonal

$\begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix}$ lower-triangular

$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ I (identity matrix)

Special matrices:

Symmetric Matrices

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

e.g.:

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

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

test

$$\begin{pmatrix} -1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \textbf{?}$$

$$A = \begin{pmatrix} 0 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix} \quad A^T = \textbf{?}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}$$

C = A – B = ?

C = A + B = ?

$$\left(\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right)^T = \textbf{?}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

C = A B = ?

C = B A = ?

$$\left\| \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\|_2 = \textbf{?}$$

(1) Transpose

Transpose: You can think of it as
- "flipping" the rows and columns

e.g. $\begin{pmatrix} a \\ b \end{pmatrix}^T = \begin{pmatrix} a & b \end{pmatrix}$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

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

(2) Matrix Addition/Subtraction

- Matrix addition/subtraction
  - Matrices must be of same size.

(2) Matrix Addition/Subtraction An Example

- If we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}$$

then we can calculate **C** = **A** + **B** by

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} 8 & 12 \\ 11 & 15 \\ 14 & 18 \end{bmatrix}$$

(2) Matrix Addition/Subtraction An Example

- Similarly, if we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}$$

then we can calculate **C** = **A** - **B** by

$$\mathbf{C} = \mathbf{A} - \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} - \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} -6 & -8 \\ -5 & -7 \\ -4 & -6 \end{bmatrix}$$

OPERATION on MATRIX

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(3) Products of Matrices

- We write the multiplication of two matrices **A** and **B** as **AB**
- This is referred to either as
  - pre-multiplying **B** by **A**

    or
  - post-multiplying **A** by **B**
- So for matrix multiplication **AB**, **A** is referred to as the *premultiplier* and **B** is referred to as the *postmultiplier*

(3) Products of Matrices

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

Condition: n = q

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

$$AB \neq BA$$

(3) Products of Matrices

- In order to multiply matrices, they must be conformable (the number of columns in the premultiplier must equal the number of rows in postmultiplier)

- Note that
  - an (m x n) x (n x p) = (m x p)
  - an (m x n) x (p x n) = cannot be done
  - a (1 x n) x (n x 1) = a scalar (1 x 1)

Products of Matrices

- If we have $A_{(3x2)}$ and $B_{(2x3)}$ then

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \text{x} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = \mathbf{C}$$

where

test

$$\begin{aligned} c_{11} &= a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ c_{12} &= a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ c_{21} &= a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ c_{22} &= a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \\ c_{31} &= a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} \\ c_{32} &= a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} \end{aligned}$$

Matrix Multiplication An Example

- If we have $\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

then $\mathbf{AB} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} x \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = \begin{bmatrix} 30 & 66 \\ 36 & 81 \\ 42 & 96 \end{bmatrix}$

where
$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} = 1(1) + 4(2) + 7(3) = 30$$
$$c_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} = 1(4) + 4(5) + 7(6) = 66$$
$$c_{21} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} = 2(1) + 5(2) + 8(3) = 36$$
$$c_{22} = a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} = 2(4) + 5(5) + 8(6) = 81$$
$$c_{31} = a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} = 3(1) + 6(2) + 9(3) = 42$$
$$c_{32} = a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} = 3(4) + 6(5) + 9(6) = 96$$

Some Properties of Matrix Multiplication

- Note that
  - Even if conformable, **AB** does not necessarily equal **BA** (i.e., matrix multiplication is *not commutative*)
  - Matrix multiplication can be extended beyond two matrices
  - matrix multiplication is *associative*, i.e., **A**(**BC**) = (**AB**)**C**

Some Properties of Matrix Multiplication

◆Multiplication and transposition

$(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$

◆Multiplication with Identity Matrix

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

Special Uses for Matrix Multiplication

- Products of Scalars & Matrices ➔ Example, If we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad b = 3.5$$

then we can calculate b**A** by

$$\mathbf{bA} = 3.5\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 3.5 & 7.0 \\ 10.5 & 14.0 \\ 17.5 & 21.0 \end{bmatrix}$$

◈Note that b**A** = **A**b if b is a scalar

Special Uses for Matrix Multiplication

- Dot (or Inner) Product of two Vectors
  - Premultiplication of a column vector **a** by conformable row vector **b** yields a single value called the *dot product* or *inner product* - If

$$\mathbf{a} = \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix}$$

then **ab** gives us

$$\mathbf{ab} = \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} = 3(5) + 4(2) + 6(8) = 71$$

which is the sum of products of elements in similar positions for the two vectors

Special Uses for Matrix Multiplication

- Outer Product of two Vectors
  - Postmultiplication of a column vector **a** by conformable row vector **b** yields a matrix containing the products of each pair of elements from the two matrices (called the *outer product*) - If

$$\mathbf{a} = \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix}$$

then **ba** gives us

$$\mathbf{ba} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 15 & 20 & 30 \\ 6 & 8 & 12 \\ 24 & 32 & 48 \end{bmatrix}$$

Special Uses for Matrix Multiplication

- Sum the Squared Elements of a Vector
  - Premultiply a column vector **a** by its transpose
    – If

$$\mathbf{a} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix}$$

then premultiplication by a row vector $\mathbf{a}^T$

$$a^T = \begin{bmatrix} 5 & 2 & 8 \end{bmatrix}$$

will yield the sum of the squared values of elements for **a**, i.e.

$$a^T\mathbf{a} = \begin{bmatrix} 5 & 2 & 8 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} = 5^2 + 2^2 + 8^2 = 93$$

Special Uses for Matrix Multiplication

- Postmultiply a row vector **a** by its transpose
- – If $\mathbf{a} = \begin{bmatrix} 7 & 10 & 1 \end{bmatrix}$

then postmultiplication by a column vector $\mathbf{a}^T$

$$a^T = \begin{bmatrix} 7 \\ 10 \\ 1 \end{bmatrix}$$

will yield the sum of the squared values of elements for **a**, i.e.

$$\mathbf{a}\, a^T = \begin{bmatrix} 7 & 10 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 10 \\ 1 \end{bmatrix} = 7^2 + 10^2 + 1^2 = 150$$

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

Yanjun Qi / UVA CS 4501-01-6501-07

(4) Vector norms

A norm of a vector ||x|| is informally a measure of the "length" of the vector.

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

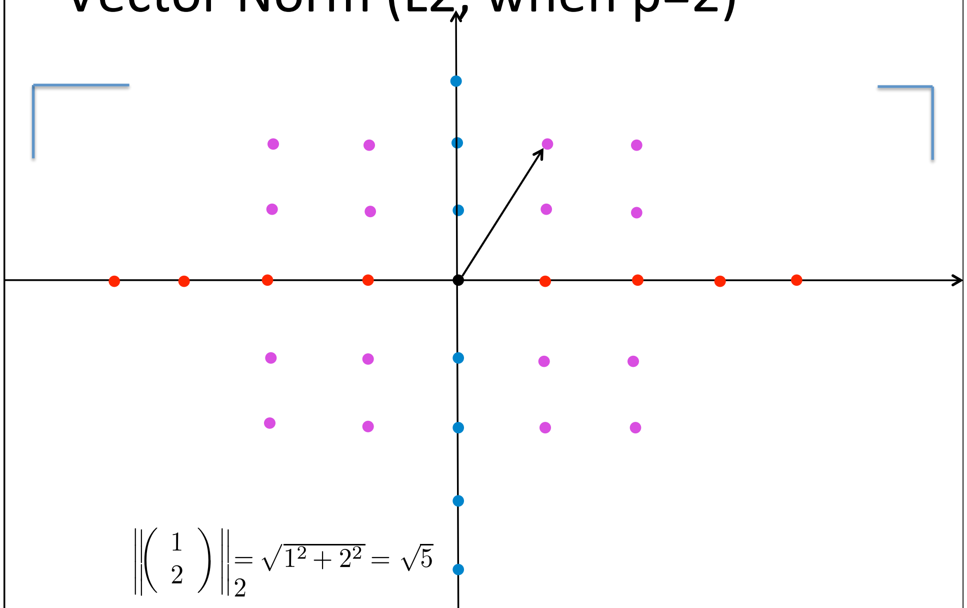
- Common norms: L_1, L_2 (Euclidean)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

- $L_{infinity}$

$$\|x\|_\infty = \max_i |x_i|$$

Vector Norm (L2, when p=2)



$$\left\| \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\|_2 = \sqrt{1^2 + 2^2} = \sqrt{5}$$

Orthogonal & Orthonormal

Inner Product defined between column vector **x** and **y** , as

➔ x•y= $x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ x_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i.$

If u•v=0, $||u||_2$!= 0, $||v||_2$!= 0

→ *u and v are orthogonal*

If u•v=0, $||u||_2$ = 1, $||v||_2$ = 1

→ *u and v are orthonormal*

Orthogonal matrices

- Notation:

$$A = \begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & . & a_{mn} \end{bmatrix}, \Rightarrow \begin{array}{l} u_1^T = [a_{11}\ a_{12} \cdots a_{1n}] \\ u_2^T = [a_{21}\ a_{22} \cdots a_{2n}] \\ \cdots \\ u_m^T = [a_{m1}\ a_{m2} \cdots a_{mn}] \end{array} \Rightarrow A = \begin{bmatrix} u_1^T \\ u_2^T \\ \cdots \\ u_m^T \end{bmatrix}$$

- A is orthogonal if:

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

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

Orthonormal matrices

- A is orthonormal if:

$$(1)\ u_k . u_k = 1 \text{ or } ||u_k|| = 1, \text{ for every } k$$

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

- Note that if A is orthonormal, it easy to find its inverse:

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

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

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(5) Inverse of a Matrix

- The inverse of a matrix **A** is commonly denoted by $\mathbf{A}^{-1}$ or inv **A**.
- The inverse of an *n x n* matrix **A** is the matrix $\mathbf{A}^{-1}$ such that $\mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$
- The matrix inverse is analogous to a scalar reciprocal
- A matrix which has an inverse is called *nonsingular*

(5) Inverse of a Matrix

- For some *n x n* matrix **A** , an inverse matrix $\mathbf{A}^{-1}$ *may not exist*.
- A matrix which does not have an inverse is singular.
- An inverse of *n x n* matrix **A** exists iff $\boxed{|\mathbf{A}| \neq 0}$

THE DETERMINANT OF A MATRIX

- The determinant of a matrix A is denoted by |**A**| (or det(**A**)).
- Determinants exist only for square matrices.
- E.g. If A = $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

$$|A| = a_{11}a_{22} - a_{12}a_{21}$$

THE DETERMINANT OF A MATRIX

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

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

THE DETERMINANT OF A MATRIX

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

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

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

HOW TO FIND INVERSE MATRIXES?
An example,

- If
- $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $|\mathbf{A}| \neq 0$

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Matrix Inverse

- The inverse A^{-1} of a matrix A has the property:
  $\boldsymbol{AA^{-1}=A^{-1}A=I}$
- A^{-1} exists only if $det(A) \neq 0$
- Terminology
  - **Singular matrix:** A^{-1} does not exist
  - **Ill-conditioned matrix:** A is close to being singular

PROPERTIES OF INVERSE MATRICES

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

Inverse of special matrix

- For diagonal matrices $\mathbf{D}^{-1} = \mathrm{diag}\{d_1^{-1}, \ldots, d_n^{-1}\}$
- For orthonormal matrices $\mathbf{A}^{-1} = \mathbf{A}^\top$
  - a square matrix with real entries whose columns and rows are orthogonal unit vectors (i.e., orthonormal vectors)

Pseudo-inverse

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

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

- It can be shown that:

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

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(6) Rank: Linear independence

- A set of vectors is linearly independent if none of them can be written as a linear combination of the others.

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

x3 = −2x1 + x2

➔ NOT linearly independent

(6) Rank: Linear independence

- **Alternative definition:** Vectors $v_1,\ldots,v_k$ are linearly independent if $c_1v_1+\ldots+c_kv_k = 0$ implies $c_1=\ldots=c_k=0$

$$\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

e.g.

$$\begin{pmatrix} 1 & 0 \\ 2 & 3 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

(u,v)=(0,0), i.e. the columns are linearly independent.

(6) Rank of a Matrix

- rank(A) (the rank of a m-by-n matrix A) is
  = The maximal number of linearly independent columns
  =The maximal number of linearly independent rows

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$$

Rank=? **Rank=?**

- If A is n by m, then
  - rank(A)<= min(m,n)
  - If n=rank(A), then A has full row rank
  - If m=rank(A), then A has full column rank

(6) Rank of a Matrix

- Equal to the dimension of the largest square sub-matrix 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

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

(6) 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

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(7) Matrix Calculus: Types of Matrix Derivatives

	Scalar	Vector	Matrix
Scalar	$\frac{\mathrm{d}y}{\mathrm{d}x}$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} = \left[\frac{\partial y_i}{\partial x}\right]$	$\frac{\mathrm{d}\mathbf{Y}}{\mathrm{d}x} = \left[\frac{\partial y_{ij}}{\partial x}\right]$
Vector	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y}{\partial x_j}\right]$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j}\right]$	
Matrix	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{X}} = \left[\frac{\partial y}{\partial x_{ji}}\right]$		

By Thomas Minka. Old and New Matrix Algebra Useful for Statistics

For Examples

$$\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T$$

$$\frac{\partial \mathbf{x}^T\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x}$$

Today Recap

- ❑ **Data Representation**
- ❑ **Linear Algebra Review**

References

- ❑ http://www.cs.cmu.edu/~zkolter/course/15-884/linalg-review.pdf (please read)
- ❑ Prof. James J. Cochran's tutorial slides "Matrix Algebra Primer II"
- ❑ http://www.cs.cmu.edu/~aarti/Class/10701/recitation/LinearAlgebra_Matlab_Rev iew.ppt
- ❑ Prof. Alexander Gray's slides
- ❑ Prof. George Bebis' slides