Machine Learning

Math Essentials

Areas of math essential to machine learning

- Machine learning is part of both ***statistics*** and computer science
  - Probability
  - Statistical inference
  - Validation
  - Estimates of error, confidence intervals
- ***Linear algebra***
  - Hugely useful for compact representation of linear transformations on data
  - Dimensionality reduction techniques
- ***Optimization*** theory

Why worry about the math?

- There are lots of easy-to-use machine learning packages out there.
- After this course, you will know how to apply several of the most general-purpose algorithms.

HOWEVER

- To get really useful results, you need good mathematical intuitions about certain general machine learning principles, as well as the inner workings of the individual algorithms.

Why worry about the math?

These intuitions will allow you to:

– Choose the right algorithm(s) for the problem
– Make good choices on parameter settings, validation strategies
– Recognize over- or underfitting
– Troubleshoot poor / ambiguous results
– Put appropriate bounds of confidence / uncertainty on results
– Do a better job of coding algorithms or incorporating them into more complex analysis pipelines

Notation

- $a \in A$ *set membership*: *a* is *member* of set *A*
- $|\ B\ |$ *cardinality*: number of items in set *B*
- $\|\ \mathbf{v}\ \|$ *norm*: length of vector *v*
- $\sum$ *summation*
- $\int$ *integral*
- $\Re$ the set of *real* numbers
- $\Re^n$ *real number space* of dimension *n*

  n = 2 : plane or 2-space
  n = 3 : 3- (dimensional) space
  n > 3 : *n*-space or *hyperspace*

Notation

- $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{u}$, $\mathbf{v}$ — *vector* (bold, lower case)
- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{X}$ — *matrix* (bold, upper case)
- $y = f(x)$ — *function* (*map*): assigns unique value in range of y to each value in domain of x
- dy / dx — *derivative* of y with respect to single variable x
- $y = f(\mathbf{x})$ — *function* on multiple variables, i.e. a vector of variables; *function* in n-space
- $\partial y / \partial x_i$ — *partial derivative* of y with respect to element i of vector $\mathbf{x}$

Linear algebra applications

1) Operations on or between vectors and matrices
2) Coordinate transformations
3) Dimensionality reduction
4) Linear regression
5) Solution of linear systems of equations
6) Many others

Applications 1) – 4) are directly relevant to this course. Today we'll start with 1).

Why vectors and matrices?

- Most common form of data organization for machine learning is a 2D array, where
  - *rows* represent samples (records, items, datapoints)
  - *columns* represent attributes (features, variables)
- Natural to think of each sample as a *vector* of attributes, and whole array as a *matrix*

vector

Refund	Marital Status	Taxable Income	Cheat
Yes	Single	125K	No
No	Married	100K	No
No	Single	70K	No
Yes	Married	120K	No
No	Divorced	95K	Yes
No	Married	60K	No
Yes	Divorced	220K	No
No	Single	85K	Yes
No	Married	75K	No
No	Single	90K	Yes

matrix

Vectors

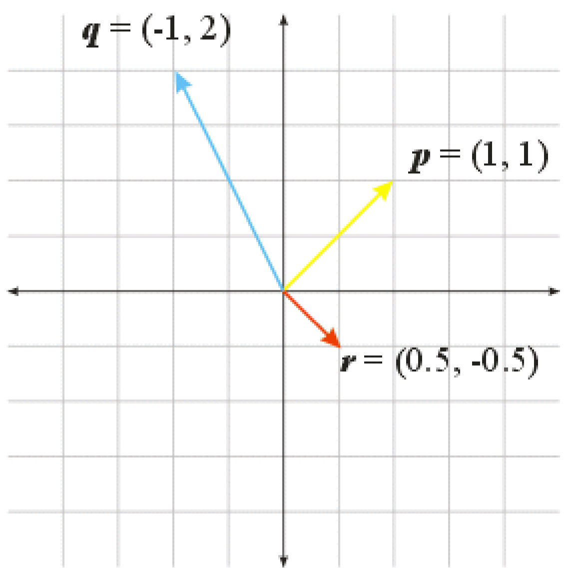
- Definition: an *n*-tuple of values (usually real numbers).
  - *n* referred to as the *dimension* of the vector
  - *n* can be any positive integer, from 1 to infinity
- Can be written in column form or row form
  - Column form is conventional
  - Vector elements referenced by subscript

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \mathbf{x}^{\mathrm{T}} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$$

$$^{\mathrm{T}}\ \text{means "transpose"}$$

Vectors

- Can think of a vector as:
  - a point in space *or*
  - a directed line segment with a magnitude and direction

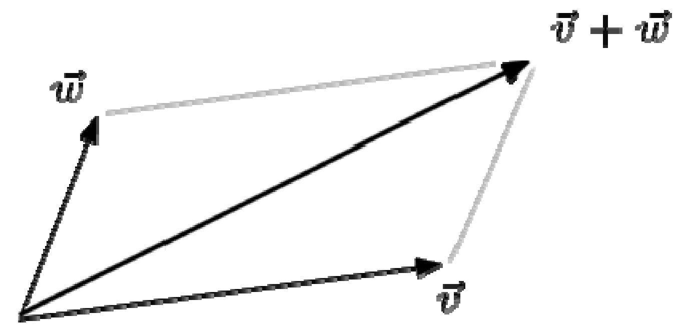


Vector arithmetic

- Addition of two vectors
  - add corresponding elements

$$\mathbf{z} = \mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 & \cdots & x_n + y_n \end{pmatrix}^{\mathrm{T}}$$

  - result is a vector



- Scalar multiplication of a vector
  - multiply each element by scalar

$$\mathbf{y} = a\mathbf{x} = \begin{pmatrix} a\,x_1 & \cdots & ax_n \end{pmatrix}^{\mathrm{T}}$$

  - result is a vector

Vector arithmetic

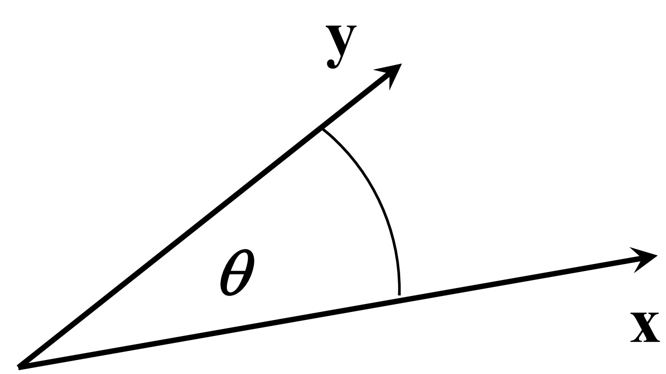
- Dot product of two vectors
  - multiply corresponding elements, then add products

$$a = \mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

  - result is a <u>scalar</u>

- Dot product alternative form

$$a = \mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \, \|\mathbf{y}\| \cos(\theta)$$



Matrices

- Definition: an *m* x *n* two-dimensional array of values (usually real numbers).
  - *m* rows
  - *n* columns
- Matrix referenced by two-element subscript
  - first element in subscript is row
  - second element in subscript is column
  - example: $\mathbf{A}_{24}$ or a_{24} is element in second row, fourth column of **A**

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

Matrices

- A vector can be regarded as special case of a matrix, where one of matrix dimensions = 1.
- Matrix *transpose* (denoted ^{T})
  - swap columns and rows
    - row 1 becomes column 1, etc.
  - *m* x *n* matrix becomes *n* x *m* matrix
  - example:

$$\mathbf{A} = \begin{pmatrix} 2 & 7 & -1 & 0 & 3 \\ 4 & 6 & -3 & 1 & 8 \end{pmatrix} \qquad \mathbf{A}^{T} = \begin{pmatrix} 2 & 4 \\ 7 & 6 \\ -1 & -3 \\ 0 & 1 \\ 3 & 8 \end{pmatrix}$$

Matrix arithmetic

- Addition of two matrices
  - matrices must be same size
  - add corresponding elements:

    $$c_{ij} = a_{ij} + b_{ij}$$

  - result is a matrix of same size

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

- Scalar multiplication of a matrix
  - multiply each element by scalar:

    $$b_{ij} = d \cdot a_{ij}$$

  - result is a matrix of same size

$$\mathbf{B} = d \cdot \mathbf{A} = \begin{pmatrix} d \cdot a_{11} & \cdots & d \cdot a_{1n} \\ \vdots & \ddots & \vdots \\ d \cdot a_{m1} & \cdots & d \cdot a_{mn} \end{pmatrix}$$

Matrix arithmetic

- Matrix-matrix multiplication
  - vector-matrix multiplication just a special case

TO THE BOARD!!

- Multiplication is associative

$$\mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{C}) = (\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{C}$$

- Multiplication is *not* commutative

$$\mathbf{A} \cdot \mathbf{B} \neq \mathbf{B} \cdot \mathbf{A} \qquad \text{(generally)}$$

- Transposition rule:

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

Matrix arithmetic

- *RULE*: In any chain of matrix multiplications, the *column* dimension of one matrix in the chain must match the *row* dimension of the *following* matrix in the chain.
- Examples

$\mathbf{A}$ 3 x 5 $\mathbf{B}$ 5 x 5 $\mathbf{C}$ 3 x 1

Right:

$\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{A}^T \quad \mathbf{C}^T \cdot \mathbf{A} \cdot \mathbf{B} \quad \mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{B} \quad \mathbf{C} \cdot \mathbf{C}^T \cdot \mathbf{A}$

Wrong:

$\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{A} \quad \mathbf{C} \cdot \mathbf{A} \cdot \mathbf{B} \quad \mathbf{A} \cdot \mathbf{A}^T \cdot \mathbf{B} \quad \mathbf{C}^T \cdot \mathbf{C} \cdot \mathbf{A}$

Vector projection

- Orthogonal projection of **y** onto **x**
  - Can take place in any space of dimensionality $\geq$ 2
  - Unit vector in direction of **x** is

    $$\mathbf{x} \,/\, \|\, \mathbf{x} \,\|$$

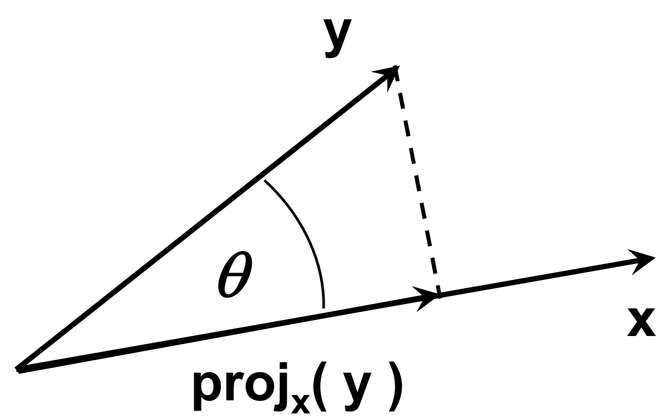
  - Length of projection of **y** in direction of **x** is

    $$\|\, \mathbf{y} \,\| \cdot \cos(\theta\,)$$

  - Orthogonal projection of **y** onto **x** is the vector

    $$\mathbf{proj_x}(\,\mathbf{y}\,) \;=\; \mathbf{x} \cdot \|\, \mathbf{y} \,\| \cdot \cos(\theta\,) \,/\, \|\, \mathbf{x} \,\| \;=$$

    $[\,(\,\mathbf{x} \cdot \mathbf{y}\,) \,/\, \|\, \mathbf{x} \,\|^2\,]\,\mathbf{x}$ (using dot product alternate form)



Optimization theory topics

- Maximum likelihood
- Expectation maximization
- Gradient descent