DSC 140A

Probabilistic Modeling & Machine Learning

Math Review #1

Math for Machine Learning

- DSC 140A is a course in **machine learning**.
- In ML, we often turn the problem of learning into a math problem.
- So, to deeply understand an ML algorithm, you need to understand the math behind it.

Math Prerequisites

- MATH 20A-B-C: Multivariate Calculus
  - Especially the **gradient**!
- MATH 18: Linear Algebra
- MATH 183: Probability / Statistics
- DSC 40A: Mathematical Foundations of ML

This Discussion

- We'll review some of the math we'll need in the first part of the course.
- It's OK to not remember everything!
- But you may want to do some review on your own.

Facts

We'll highlight some important facts throughout this discussion with a box like this:

Fact #1

This is a fact.

DSC 140A

Probabilistic Modeling & Machine Learning

Summation Notation

Summation Notation

- We use summation notation a lot in data science.

- If $x_1, x_2, \ldots, x_n$ are numbers (or vectors), then:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

Exercise

True or False: constant factors can be pulled out of a summation. That is, if a is a constant (indepedent of i), then:

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$$

True!

Fact #2 Constant Factors in a Summation

Constants can be pulled out of a summation. That is, if a is a constant (independent of i), then:

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$$

How do we know?

- Try expanding the sum using … notation:

$$\begin{aligned}\sum_{i=1}^{n} ax_i &= ax_1 + ax_2 + \ldots + ax_n \\ &= a(x_1 + x_2 + \ldots + x_n) \\ &= a\sum_{i=1}^{n} x_i\end{aligned}$$

Exercise

True or False: we can “split” a summation. That is:

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

True!

Fact #3 Splitting a Summation

We can "split" a summation. That is:

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

Exercise

How should we interpret $\sum_{i=1}^{n} x_i + y_i$

$$\sum_{i=1}^{n} (x_i + y_i) \quad \text{or} \quad \left(\sum_{i=1}^{n} x_i\right) + y_i$$

Answer

- It *has* to mean $\sum_{i=1}^{n}(x_i + y_i)$, because $\left(\sum_{i=1}^{n} x_i\right) + y_i$ does not make notational sense!
- i is “unbound”, so y_i is not defined!

DSC 140A

Probabilistic Modeling & Machine Learning

Vectors

Vectors

- A **vector** $\vec{x}$ is a list of numbers.
- The **dimensionality** of the vector is the number of entries it has.
- Example: a 3-vector:

$$\vec{x} = \begin{pmatrix} 8 \\ -2 \\ 3 \end{pmatrix}$$

Vector Notation

- We write $x \in \mathbb{R}^d$, to denote that $\vec{x}$ is a d-dimensional vector whose entries are real numbers.[1] [2]

- Pronounced "x is in R-d".

---

[1] $\mathbb{R}$ is the symbol for the set of real numbers.

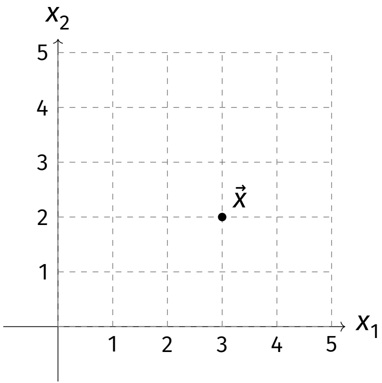
[2] In LaTeX, you can write `\vec{x} \in\mathbb R^d`

Vector Notation

- We use subscripts to denote particular elements of a vector.
- Example: x_1 is the first element of $\vec{x}$, x_2 is the section element, etc.

Points vs. Arrows

- We often think of vectors as **points** in space.
  - Example: $\vec{x} = (3, 2)^T$

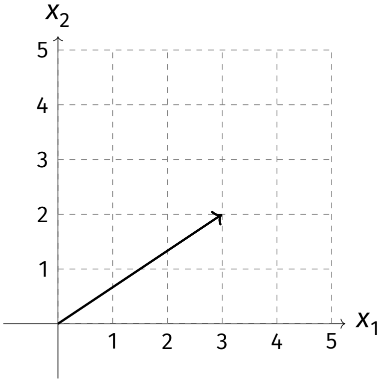


Vector Notation

- We'll often be working with sets of vectors.
- We'll use a superscript to denote the ith vector in the set.
- $\vec{x}^{(1)}$ is the first vector in the set, $\vec{x}^{(2)}$ is the second, etc.

Points vs. Arrows

- We can also think of vectors as arrows.
  - Example: $\vec{x} = (3, 2)^T$



Vector Norm (Length)

- The **norm** (length) of a vector $\vec{x}$, written $\|\vec{x}\|$, is the Euclidean distance from the origin to the point represented by $\vec{x}$:

$$\|\vec{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_d^2}$$
$$= \sqrt{\sum_{i=1}^{d} x_i^2}$$

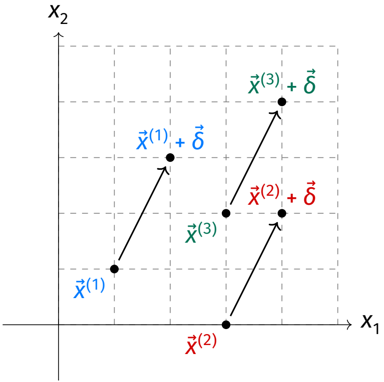
Vector Addition

- Two vectors $\vec{x}$ and $\vec{y}$ can be added together.
- The result is a vectors whose entries are the *elementwise* sum of the two vectors.
- Example:

$$\underbrace{\begin{pmatrix}1\\2\\3\end{pmatrix}}_{\vec{x}} + \underbrace{\begin{pmatrix}4\\5\\6\end{pmatrix}}_{\vec{y}} = \begin{pmatrix}1+4\\2+5\\3+6\end{pmatrix} = \underbrace{\begin{pmatrix}5\\7\\9\end{pmatrix}}_{\vec{x}+\vec{y}}$$

Fact #4 Vector Addition

Adding (or subtracting) $\vec{\delta}$ to $\vec{x}$ "shifts" $\vec{x}$. For example, using $\vec{\delta} = (1, 2)^T$:



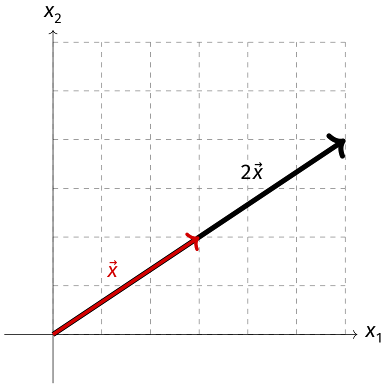
Scalar Multiplication

- We can multiply a vector by a scalar, c.
- The result is a vector whose entries are the original entries multiplied by c.
- Example:

$$3\begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}3\cdot 1\\3\cdot 2\\3\cdot 3\end{pmatrix} = \begin{pmatrix}3\\6\\9\end{pmatrix}$$

Fact #5 Scalar Multiplication of a Vector

Multiplying $\vec{x}$ by c "stretches" $\vec{x}$ by a factor of c. For example, using $c = 2$:



Vector Products

- We can “multiply” two vectors together using the **dot product**.

Fact #6 Dot Product (Coordinate Definition)

The **dot product** of two d-vectors $\vec{u}$ and $\vec{v}$ is defined to be:

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \ldots + u_d v_d$$
$$= \sum_{i=1}^{d} u_i v_i$$

Exercise

Let $\vec{u} = (1, 2, 3)^T$ and $\vec{v} = (4, 5, 6)^T$. What is $\vec{u} \cdot \vec{v}$?

Dot Product

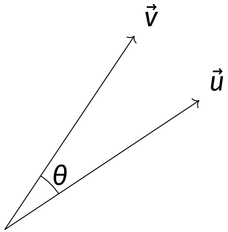
- The dot product has a geometric interpretation, too.

Fact #7 Dot Product (Geometric Definition)

The dot product of two vectors $\vec{u}$ and $\vec{v}$ is:

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \, \|\vec{v}\| \cos\theta$$

where θ is the angle between the two vectors.



Exercise

Suppose $\vec{v} = (3, 3)^T$.

1. Find a unit vector $\vec{u}^{(1)}$ such that $\vec{u}^{(1)} \cdot \vec{v} = 0$.

2. Find a unit vector $\vec{u}^{(2)}$ such that $|\vec{u}^{(2)} \cdot \vec{v}|$ is maximized.

Exercise

Which of these is another expression for the norm of $\vec{u}$?

a) $\vec{u} \cdot \vec{u}$

b) $\sqrt{\vec{u}^2}$

c) $\sqrt{\vec{u} \cdot \vec{u}}$

d) $\vec{u}^2$

Fact #8 Properties of the Dot Product

The dot product is:

- Commutative: $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
- Distributive: $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$
- Linear: $\vec{u} \cdot (\alpha\vec{v} + \beta\vec{w}) = \alpha\vec{u} \cdot v + \beta\vec{u} \cdot \vec{w}$

DSC 140A

Probabilistic Modeling & Machine Learning

Matrices

Matrices

An $m \times n$ **matrix** is a table of numbers with m rows, n columns:

- Example: 2×3 matrix:

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{pmatrix}$$

Matrices

An $m \times n$ **matrix** is a table of numbers with m rows, n columns:

- Example: 3×3 "square" matrix:

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Matrices

An $m \times n$ **matrix** is a table of numbers with m rows, n columns:

- Example: 3×1, a.k.a. a "column vector":

$$\begin{pmatrix} m_{11} \\ m_{21} \\ m_{31} \end{pmatrix}$$

Matrix Notation

- We use upper-case letters for matrices.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

- Sometimes use subscripts to denote particular elements: $A_{13} = 3$, $A_{21} = 4$

Matrix Transpose

- A^T denotes the **transpose** of A:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Matrix Addition and Scalar Multiplication

- We can add two matrices...
- But **only** if they are the same shape!
- Addition occurs elementwise:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 7 & 8 & 9 \\ -1 & -2 & -3 \end{pmatrix} = \begin{pmatrix} 8 & 10 & 12 \\ 3 & 3 & 3 \end{pmatrix}$$

Scalar Multiplication

- Scalar multiplication occurs elementwise, too:

$$2 \cdot \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix}$$

Matrix-Vector Multiplication

- We can multiply an $m \times n$ matrix A by an n-vector $\vec{x}$...
- Note that the number of columns in A **must** equal the number of entries in $\vec{x}$!
- The result is an m-vector.

Fact #9 Matrix-Vector Mult., View 1

Let A be an $m \times n$ matrix and $\vec{x}$ be an n-vector.

The ith entry of $A\vec{x}$ can be found by dotting the ith row of A with $\vec{x}$.

Example

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

$$(A\vec{x})_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = 3 + 4 + 1 = 8$$

$$(A\vec{x})_2 = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = 9 + 8 + 5 = 22$$

$$A\vec{x} = (8, 22)^T$$

Fact #10 Matrix-Vector Mult., View 2

Let A be an $m \times n$ matrix and $\vec{x} = (x_1, \ldots, x_n)$ be an n-vector.

$A\vec{x}$ equals:

- x_1 times the first column of A, plus
- x_2 times the second column of A, plus
- ..., plus
- x_n times the nth column of A.

Example

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

$$\begin{aligned} A\vec{x} &= 3\begin{pmatrix} 1 \\ 3 \end{pmatrix} + 2\begin{pmatrix} 2 \\ 4 \end{pmatrix} + 1\begin{pmatrix} 1 \\ 5 \end{pmatrix} \\ &= \begin{pmatrix} 3 \\ 9 \end{pmatrix} + \begin{pmatrix} 4 \\ 8 \end{pmatrix} + \begin{pmatrix} 1 \\ 5 \end{pmatrix} \\ &= \begin{pmatrix} 8 \\ 22 \end{pmatrix} \end{aligned}$$

Fact #11 Matrix-Vector Mult., View 3

Let A be an $m \times n$ matrix and $\vec{x}$ be an n-vector.
The ith entry of $A\vec{x}$ is given by:

$$\sum_{j=1}^{n} A_{ij} x_j$$

Matrix-Matrix Multiplication

- We can multiply two matrices A and B if (and only if) # cols in A is equal to # rows in B
- If $A = m \times n$ and $B = n \times p$, the result is $m \times p$.
  - This is **very useful**. Remember it!

Fact #12 Matrix-Matrix Mult., View 1

Let A be an $m \times n$ matrix and B be an $n \times p$ matrix.

The (i, j)th entry of AB is given by dotting the ith row of A with the jth column of B.

Fact #13 Matrix-Matrix Mult., View 2

Let A be an $m \times n$ matrix and B be an $n \times p$ matrix.

The (i, j)th entry of AB is given by:

$$\sum_{k=1}^{n} A_{ik} B_{kj}$$

Matrix-Matrix Multiplication Example

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 3 & 6 \\ 1 & 3 \\ 4 & 8 \end{pmatrix}$$

- What is the size of AB?
- What is $(AB)_{12}$?

Fact #14 Matrix Multiplication Properties

Matrix multiplication is:

- Distributive: $A(B + C) = AB + AC$
- Associative: $(AB)C = A(BC)$
- **Not commutative in general**: $AB \neq BA$

Fact #15 $\vec{u} \cdot \vec{v}$ as Matrix Multiplication

An n-vector can be thought of an an $(n \times 1)$ matrix. So the dot product of two n-vectors $\vec{u}$ and $\vec{v}$ is the same as the matrix multiplication $\vec{u}^T\vec{v}$.

Identity Matrices

- The $n \times n$ **identity matrix** I has ones along the diagonal:
$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$
- If A is $n \times m$, then $IA = A$.
- If B is $m \times n$, then $BI = B$.

Systems of Linear Equations

- We often want to solve $A\vec{x} = \vec{b}$ for $\vec{x}$.
- There are three possible situations:
  1. There's no solution.
  2. There's exactly one solution.
  3. There are infinitely many solutions.

Solving Systems

- If A is $n \times n$, then it might have an **inverse**.
- The inverse of A, denoted A^{-1}, is the matrix such that $AA^{-1} = I$.
- The inverse, if it exists, is also $n \times n$.

Fact #16 Matrix Inverse

Suppose A is $n \times n$, and we want to solve $A\vec{x} = \vec{b}$ for $\vec{x}$.

If A is **invertible** (has an inverse), then there is a unique solution: $\vec{x} = A^{-1}\vec{b}$.

If A is **not invertible** then there is either no solution or infinitely many solutions.

Matrix Inverse

- You don’t know how to compute matrix inverses by hand for this class.
- But you do need to know these properties.

DSC 140A

Probabilistic Modeling & Machine Learning

What kind of object?

Debugging for ML

- In this class, you'll find yourself doing some long calculations with matrices and vectors.
- It's easy to get lost in the weeds.
- It is helpful to frequently stop and ask yourself:
    1. "What kind of object *should* this be? A scalar, vector, or matrix?"
    2. "What type of object is it actually?"
- This can help you debug your ML code, too!

What kind of object?

- To answer this, remember:
  - scalar × vector → vector
  - vector + vector → vector
  - matrix + matrix → matrix
  - vector · vector (dot product) → scalar
  - vector norm → scalar
  - $(m \times n)$ matrix × n-vector → m-vector
  - $(m \times n)$ matrix × $(n \times p)$ matrix → $(m \times p)$ matrix
  - ...

Watch out for...

- The following are **not mathematically valid**.
  Make sure your calculations don't lead to these:
  - vector + scalar
  - matrix + scalar
  - matrix + vector
  - $(m \times n)$ matrix $\times$ $(p \times q)$ matrix, with $n \neq p$

Example

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ be d-dimensional vectors, and $\vec{w}$ be a d-dimensional vector. Let $y_1, \ldots, y_n$ be scalars.

What type of object is

$$\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i \cdot \vec{w} - y_i)^2$$

Example

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ be d-dimensional vectors, and $\vec{w}$ be a d-dimensional vector. Let $y_1, \ldots, y_n$ be scalars.

What type of object is

$$\frac{1}{n} \sum_{i=1}^{n} (\underbrace{\vec{x}_i \cdot \vec{w}}_{\text{scalar}} - y_i)^2$$

Example

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ be d-dimensional vectors, and $\vec{w}$ be a d-dimensional vector. Let $y_1, \ldots, y_n$ be scalars.

What type of object is

$$\frac{1}{n} \sum_{i=1}^{n} (\underbrace{\vec{x}_i \cdot \vec{w} - y_i}_{\text{scalar}})^2$$

Example

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ be d-dimensional vectors, and $\vec{w}$ be a d-dimensional vector. Let $y_1, \ldots, y_n$ be scalars.

What type of object is

$$\frac{1}{n} \sum_{i=1}^{n} \underbrace{(\vec{x}_i \cdot \vec{w} - y_i)^2}_{\text{scalar}}$$

Example

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ be d-dimensional vectors, and $\vec{w}$ be a d-dimensional vector. Let $y_1, \ldots, y_n$ be scalars.

What type of object is

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} (\vec{x}_i \cdot \vec{w} - y_i)^2}_{\text{scalar}}$$

Example

Let $\vec{x}^{(1)}, \dots, \vec{x}^{(n)}$ be d-dimensional vectors, and $\vec{w}$ be a d-dimensional vector. Let $y_1, \dots, y_n$ be scalars.

What type of object is

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i \cdot \vec{w} - y_i)^2}_{\textbf{scalar}}$$

Exercise

Let $\vec{x} \in \mathbb{R}^d$ and let A be a $d \times d$ matrix. What type of object is $\vec{x}^T A \vec{x}$?

Exercise

Let $\vec{x} \in \mathbb{R}^d$ and let A be a $d \times d$ matrix. What type of object is $\vec{x}^T A \vec{x}$?

Answer: A scalar.

Exercise

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ be d-dimensional vectors. What type of object is:

$$\frac{1}{n} \sum_{i=1}^{n} \vec{x}^{(i)} (\vec{x}^{(i)})^T$$

Exercise

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ be d-dimensional vectors. What type of object is:

$$\frac{1}{n} \sum_{i=1}^{n} \vec{x}^{(i)} (\vec{x}^{(i)})^T$$

Answer: A $d \times d$ matrix.