Preliminaries

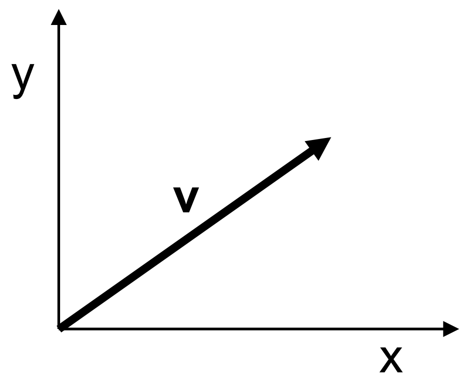
Outline

- Linear Algebra
- Calculus
- Probability

Vector

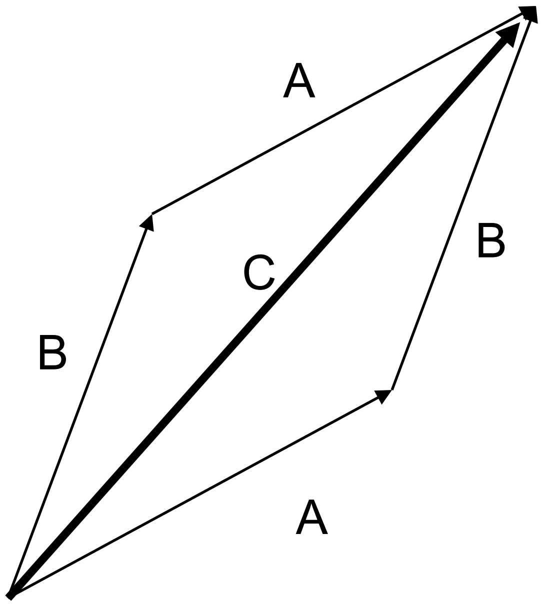
- Think of a vector as a *<u>directed line segment in N-dimensions</u>*! (has "length" and "direction")
- Basic idea: convert geometry in higher dimensions into algebra!
  - Once you define a "nice" *<u>basis</u>* along each dimension: x-, y-, z-axis …
  - Vector becomes a N x 1 matrix!
- 1-dimensional array

$$\vec{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$



Vector Addition: A+B

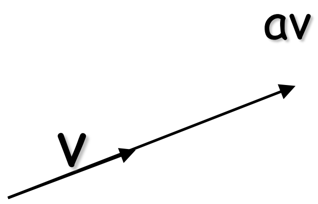
$$\mathbf{A+B} = (x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$$



A+B = C
(use the head-to-tail method to combine vectors)

Scalar Product: $a\mathbf{v}$

$$a\mathbf{v} = a(x_1, x_2) = (ax_1, ax_2)$$



Change only the length ("scaling"), but keep *direction fixed*.

Sneak peek: matrix operation (**Av**) can change *length, direction and also dimensionality*!

Vectors: Dot/Inner Product

$$A \cdot B = A^T B \ = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} d \\ e \\ f \end{bmatrix} = ad + be + cf$$

Think of the dot product as a matrix multiplication

$$\|A\|^2 = A^T A = aa + bb + cc$$

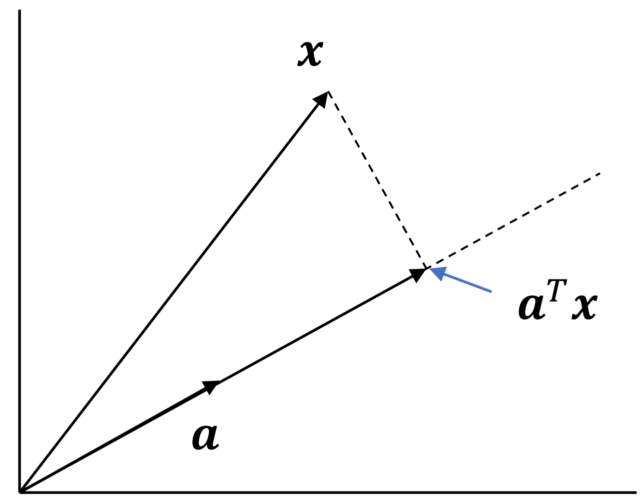
The magnitude is the dot product of a vector with itself

$$A \cdot B = \|A\| \, \|B\| \cos(\theta)$$

The dot product is also related to the angle between the two vectors

$$A \cdot B = 0 \iff A \perp B$$

Projection: Using Inner Products

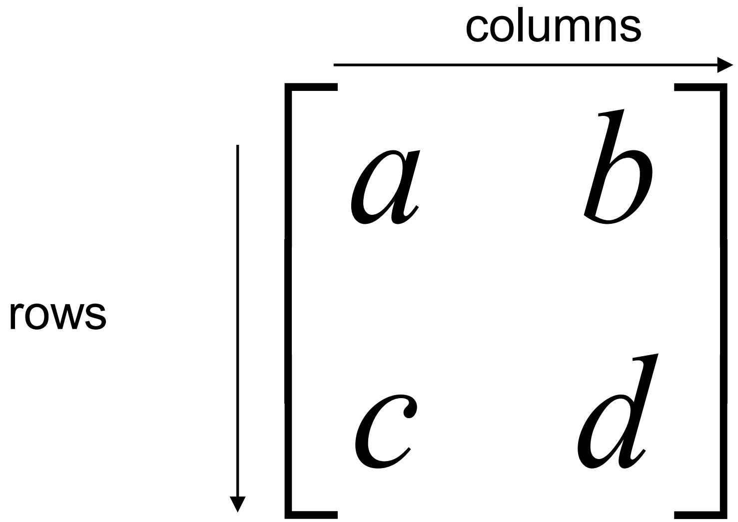


Projection of $\boldsymbol{x}$ along the direction $\boldsymbol{a}$ ($\|\boldsymbol{a}\| = 1$)

$$\mathbf{p} = \mathbf{a}\,(a^T x)$$
$$\|a\| = a^T a = 1$$

Matrix

- A matrix is a set of elements, organized into rows and columns
- $N \times M$ matrix
- 2-dimensional array
- Transpose



Elementwise Matrix Operations

- Addition, Subtraction, Multiplication: creating new matrices (or functions)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a-e & b-f \\ c-g & d-h \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae & bf \\ cg & dh \end{bmatrix}$$

Matrix Times Matrix

$$\mathbf{L} = \mathbf{M} \cdot \mathbf{N}$$

$$\begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \cdot \begin{bmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \end{bmatrix}$$

$$l_{12} = m_{11}n_{12} + m_{12}n_{22} + m_{13}n_{32}$$

Multiplication

- Is AB = BA? Maybe, but maybe not!

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & \ldots \\ \ldots & \ldots \end{bmatrix} \quad \begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ea+fc & \ldots \\ \ldots & \ldots \end{bmatrix}$$

- Matrix multiplication AB: apply transformation B first, and then again transform using A!
- Heads up: multiplication is NOT commutative!

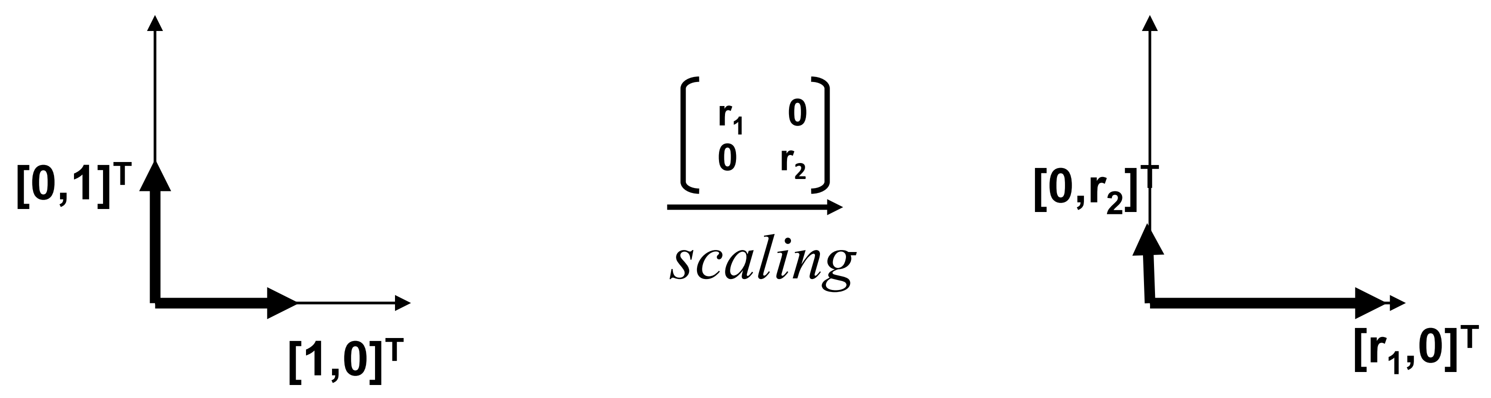
Matrix operating on vectors

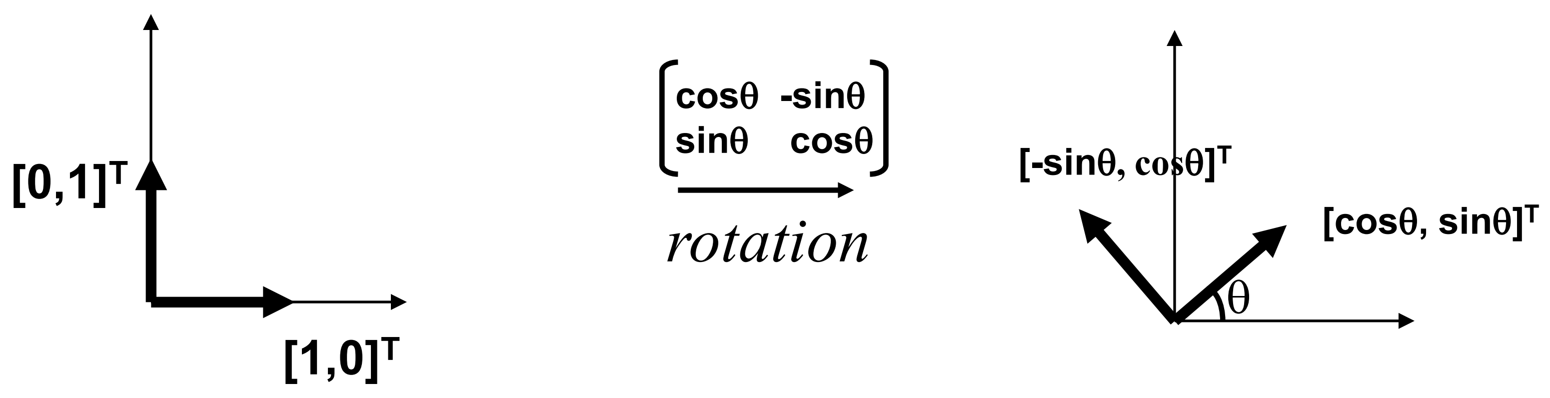
- Matrix is like a function that transforms the vectors on a plane
- Matrix operating on a general point => transforms x- and y-components
- *System of linear equations*: matrix is just the bunch of coeffs !

- x’ = ax + by
- y’ = cx + dy

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix}$$

Matrices: Scaling, Rotation





Inverse of a Matrix

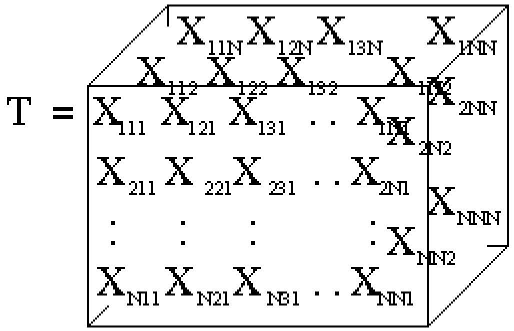
- Identity matrix:
  AI = A
- Inverse exists only for <u>square matrices</u> that are <u>non-singular</u>
- Some matrices have an inverse, such that:
  $\mathbf{AA^{-1} = I}$

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Tensors

- A generic way of describing N-dimensional arrays
  - Vector: first-order tensor
  - Matrix: second-order tensor

A three-order tensor



Derivatives and Differentiation

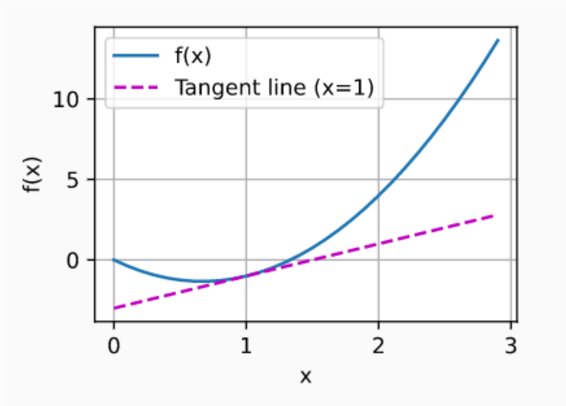
- For a function $y = f(x)$, the derivative of f is defined as

$$f'(x) = \frac{dy}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

If $f'(a)$ exists, f is said to be differentiable at a, where the derivative is $f'(a)$

- Example: $f(x) = 3x^2 - 4x$

- $f'(a)$ can also be interpreted as the slope of the tangent line to the curve of f at point a



Partial Derivatives

- Extend the ideas of differentiation to *multivariate* functions.
- Let $y = f(x_1, x_2, \ldots, x_n)$ be a function with n variables. The partial derivative of y with respect to its ith parameter x_i is

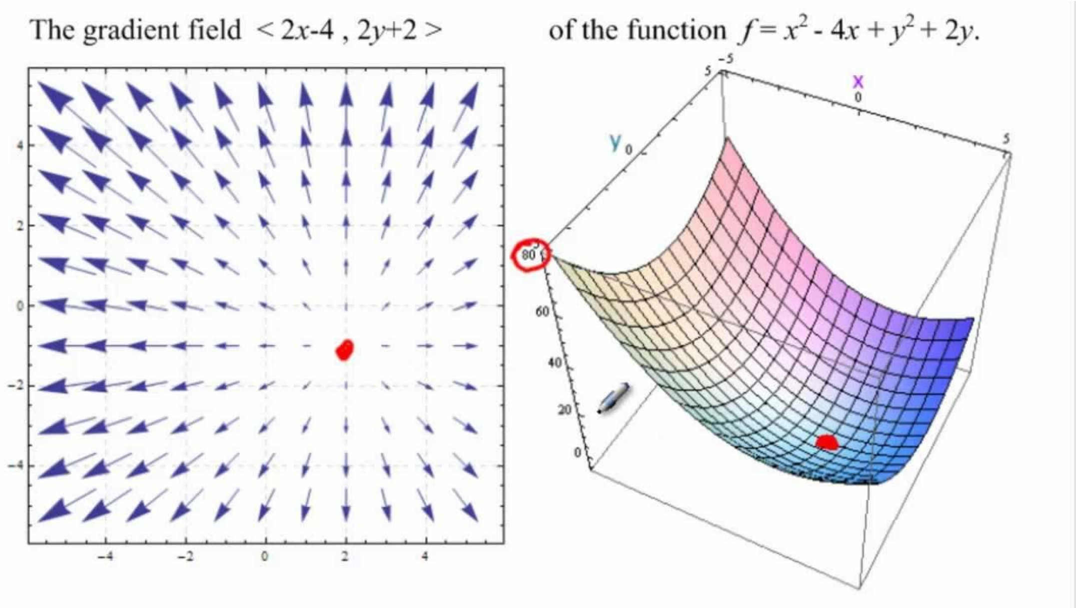
$$\frac{\partial y}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}$$

- To calculate $\frac{\partial y}{\partial x_i}$, we can simply treat $x_1, \ldots, x_i - 1, x_i + 1, \ldots, x_n$ as constants and calculate the derivative of y with respect to x_i.
- Example: $f(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2$

Gradients

- Let $\boldsymbol{x} = [x_1, x_2, \dots, x_n]$, the gradient of function $f(\boldsymbol{x})$ w.r.t. $\boldsymbol{x}$ is

$$\nabla f(\boldsymbol{x}) = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \dots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right]$$



Chain Rule

- Help us to compute derivatives for composite functions.
- Three variables: z, y, x.
  - $z = f(y), y = g(x)$

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = f'(g(x))g'(x)$$

- Extend to partial derivatives
  - $z = f(y_1, y_2, \ldots, y_m), y_i = g_i(x_1, x_2, \ldots, x_n)$

$$\frac{\partial z}{\partial x_j} = \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x_j} + \cdots + \frac{\partial z}{\partial y_m}\frac{\partial y_m}{\partial x_j}$$

Random Variables

- A variable whose value is not deterministic.
- A discrete random variable X takes value from a sample space (e.g., $S = \{1,2,3,4,5,6\}$). The distribution $P(X)$ tells us the probability that X takes any value.
- A continues random variable X takes value from a continuous domain (e.g., $\mathbb{R}$). The probability density function $f(x)$ tells us the likelihood that we see a value. The cumulative distribution function $P(x)$ tells us the probability that X will take a value less than or equal to x.

Bayes' Theorem

- Joint probability $P(A = a, B = b)$: The probability that $A = a$ and $B = b$ happen simultaneously.
- Conditional probability $P(B = b|A = a) = \frac{P(A=a,B=b)}{P(A=a)}$: The probability of $B = b$, provided that $A = a$ has occurred.
- Marginalization: $P(B) = \sum_A P(A, B)$
- Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Expectation and Variance

- The average of the random variable X is quantified by its expectation:

$$E[X] = \sum_x xP(X = x)$$

- The expectation of function $f(x)$:

$$E_{x \sim P(X)}[f(x)] = \sum_x f(x)P(x)$$

- How much the random variable X deviates from its expectation is quantified by the variance:

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$