Lecture 8: Useful Math and Linear Algebra

COMP 411, Fall 2021
Victoria Manfredi

WESLEYAN UNIVERSITY

Acknowledgements: These slides are based primarily on those created by Michael Paul (U of Colorado Boulder) and lectures by Balaraman Ravindran (IIT Madras)

Today's Topics

What happens when we have more than 2 dimensions?

Vectors and data points

- What does a feature vector look like geometrically?
- How to calculate the distance between points?
- Definitions: vector products and linear functions

vumanfredi@wesleyan.edu

Useful Math

LINEAR FUNCTIONS

Goals

So far focused on visualizations in 2-dimensions

- What happens when we have more than 2 dimensions?

Vectors and data points

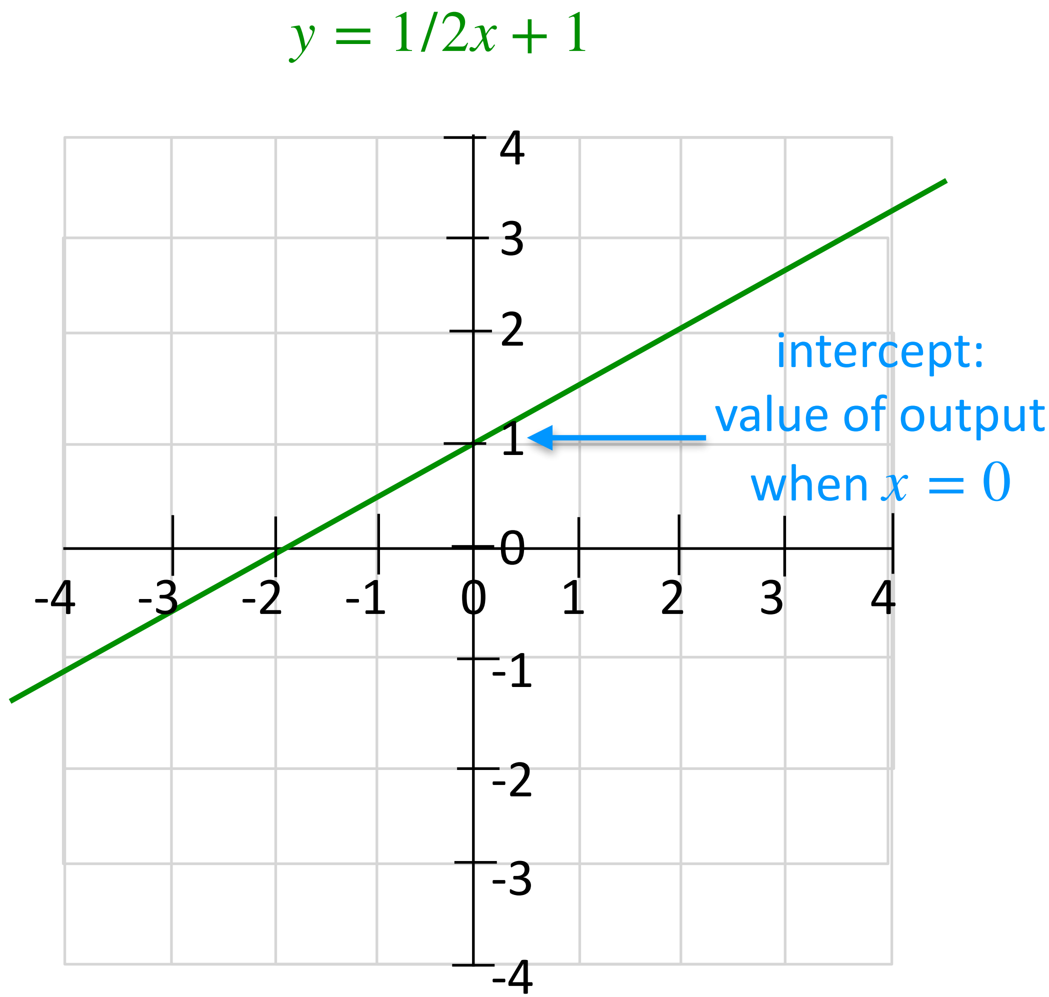
- What does a feature vector look like geometrically?
- How to calculate the distance between points?
- Definitions: linear functions and vector products

Linear functions

General form of a line:

$$f(x) = mx + b$$

slope — m; intercept — b

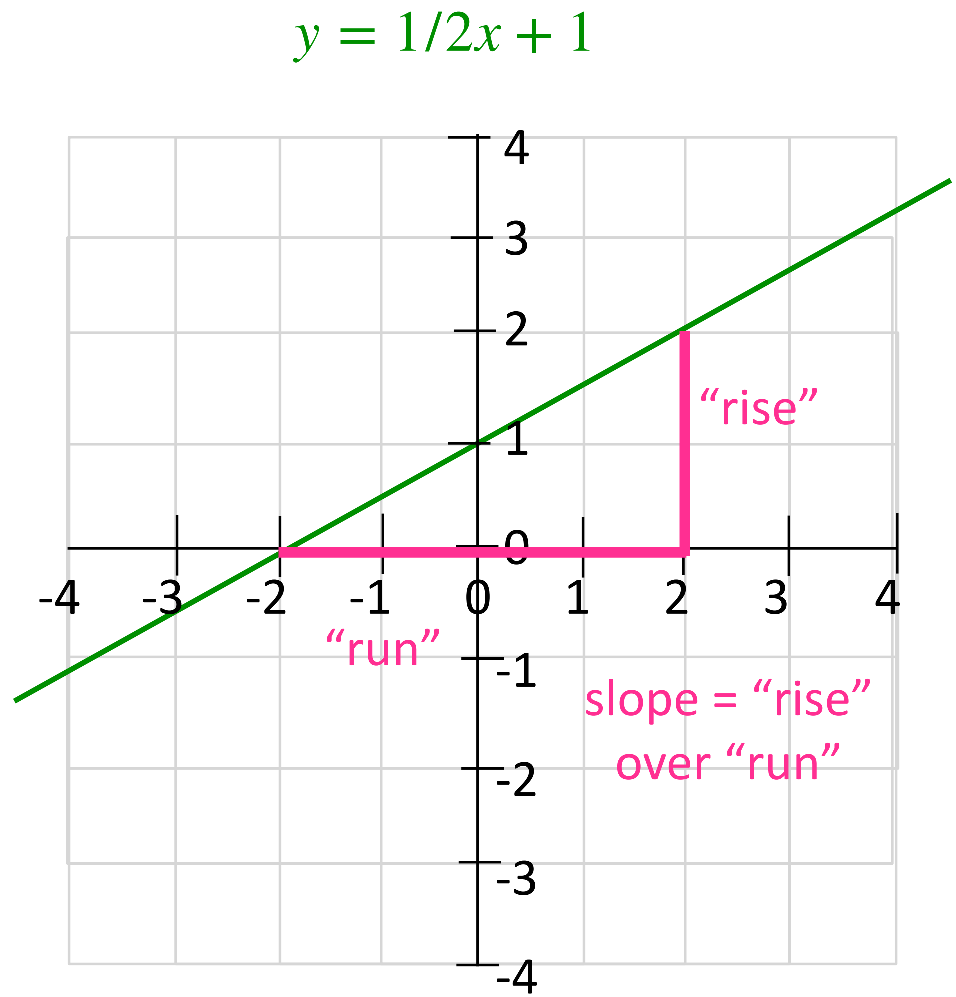


Linear functions

General form of a line:

$$f(x) = mx + b$$

slope (m) intercept (b)



Linear functions

General form of a line:

$$f(x) = mx + b$$

slope (m) intercept (b)

m and b are called **parameters** or **coefficients.** They are constant (once specified)

x is the **argument** or input of the function

Machine learning involves learning the parameters of the predictor function.

In linear regression, the predictor function is a linear function, but the parameters are unknown ahead of time. Goal is to learn what slope and intercept should be.

Linear functions

Can have more than one argument

One variable: line

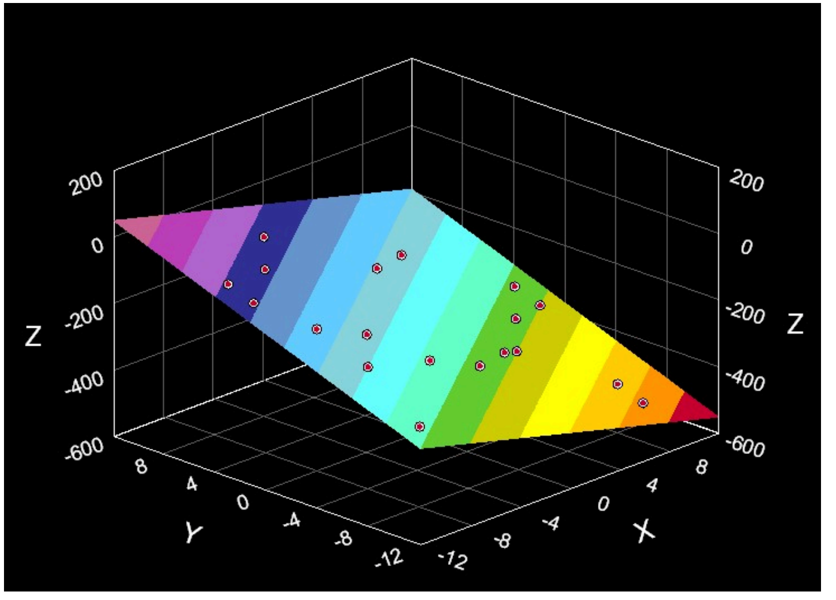
$$f(x) = mx + b$$

Two variables: plane

$$f(x_1, x_2) = m_1 x_1 + m_2 x_2 + b$$

General form: hyperplane

$$f(x_1, \ldots, x_k) = \sum_{i=1}^{k} m_i x_i + b$$



http://www.nlreg.com/plane3d.htm

Standard form of a line

$$Ax + By = C$$

A, B, and C are real numbers

A and B are not both zero

Drawing line

$\mathbf{w} \cdot \mathbf{x} + b$

2-dimensions: $w_1 x_1 + w_2 x_2 + b = 0$

Solve for x_1-intercept: $x_1 = \frac{-(b - w_2 x_2)}{w_1}$ if $y = 0$ then $x_1 = \frac{-b}{w_1}$

Solve for x_2-intercept: $x_2 = \frac{-(b - w_1 x_1)}{w_2}$ if $y = 0$ then $x_2 = \frac{-b}{w_2}$

Two points: $(0, -b/w_2)$, $(-b/w_1, 0)$

Slope $= \frac{-b/w_2}{b/w_1}$, intercept $x_2 = \frac{-b}{w_2}$

Linear algebra

VECTORS

Vector notation

A list of values is called a vector

We can use variables to denote entire vectors as shorthand

$$\mathbf{m} = < m_1, m_2, m_3, m_4 >$$

$$\mathbf{x} = < x_1, x_2, x_3, x_4 >$$

Dot product

The dot product of two vectors is written as $\mathbf{m}^T\mathbf{x}$ or $\mathbf{m} \cdot \mathbf{x}$, which is defined as: $\mathbf{m}^T\mathbf{x} = \sum_{i=1}^{k} m_i x_i$

Example

$$\mathbf{m} = < 5.13, 1.08, -0.03, 7.29 >$$

$$\mathbf{x} = < x_1, x_2, x_3, x_4 >$$

$$\mathbf{m}^T\mathbf{x} = 5.13x_1 + 1.08x_2 - 0.03x_3 + 7.29x_4$$

Vector notation

Equivalent notation for a linear function:

$$f(x_1, \ldots, x_k) = \sum_{i=1}^{k} m_i x_i + b$$

OR

$$f(\mathbf{x}) = \mathbf{m}^T \mathbf{x} + b$$

Vector notation

Terminology:

A **point** is the same as a **vector** (at least as used in this class)

Remember:

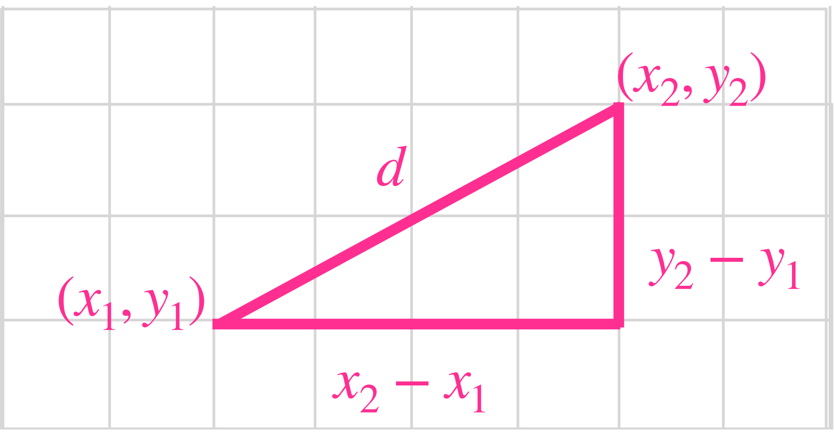
In machine learning, the number of dimensions in your points/vectors is the number of **features**

Useful Math

DISTANCE

Distance

How far apart are two points?



Euclidean distance between 2 points in 2 dimensions:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

In 3 dimensions (x, y, z):

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Distance

General formulation Euclidean distance between 2 points with k dimensions:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{k} (p_i - q_i)^2}$$

where $\mathbf{p}$ and $\mathbf{q}$ are 2 points (each represents a k -dimensional vector)

Distance

Example:

$$\mathbf{p} = < 1.3,\ 5.0,\ -0.5,\ -1.8 >$$

$$\mathbf{q} = < 1.8,\ 5.0,\ 0.1,\ -2.3 >$$

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(1.3 - 1.8)^2 + (5.0 - 5.0)^2 + (-0.5 - 0.1)^2 + (-1.8 - 2.3)^2}$$

$$= \sqrt{0.86} = 0.927$$

Distance

A special case is the distance between a point and 0 (the origin)

$$d(\mathbf{p}, \mathbf{0}) = \sqrt{\sum_{i=1}^{k} (p_i)^2}$$

This is called the Euclidean norm of $\mathbf{p}$

- A norm is a measure of a vector's length
- The Euclidean norm is also called the L2 norm
  - We'll learn about other norms later

Useful Math

LINEAR ALGEBRA

Overview

What is linear algebra? Branch of mathematics concerning vector spaces and linear mappings between such spaces. it is the study of lines, planes, and subspaces, but is also concerned with properties common to all vector spaces

Why do we study linear algebra? Provides a way to compactly represent and operate on sets of linear equations. Why do linear equations pop up in machine learning? In machine learning, we represent data as matrices and parameters as vectors and hence it is natural to use the notations and formalisms developed in linear algebra

Represent data as nxp matrix where n is # of points, p is # of features

Introduction to linear algebra

Consider the following system of equations

$$4x_1 - 5x_2 = -13$$
$$-2x_1 + 3x_2 = 9$$

2 equations and 2 variables

In matrix notation, the system is more compactly represented as

$$\mathbf{Ax} = \mathbf{b}$$
$$\mathbf{A} = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}$$
$$\mathbf{b} = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

b represents the outages

Machine learning uses linear algebra to manipulate several equations at once in multiple variables

Introduction to linear algebra

Solving without matrices? first solve for 1 variable and substitute to get other variable

Using matrices you can solve directly: just multiply both sides by $\mathbf{A}^{-}$ inverse

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\mathbf{A}^{-}\mathbf{A}\mathbf{x} = \mathbf{A}^{-}\mathbf{b}$$

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{b}$$

Of course you will have to care about all matrices don’t have an inverse, but in most cases they do; in that case ou can directly get the solution of x, solving it just through 1 equation.

Vector space

Definition: A set V with 2 operations $+$ and $\cdot$ is said to be a vector space if it is closed under both the vector addition and scalar multiplication operations and satisfies the following 8 axioms.

1. Commutative law: $x + y = y + x, \ \forall x, y \in V$
2. Associative law:
   $(x + y) + z = x + (y + z), \ \forall x, y, z \in V$
3. Additive identity: $\exists \varnothing \in V$ s. t. $x + \varnothing = x, \ \forall x \in V$
4. Additive inverse: $\forall x \in V, \exists \tilde{x}$ s. t. $x + \tilde{x} = 0$