APS1070

Foundations of Data Analytics and Machine Learning

Winter 2021

Week 5:

- *Linear Algebra*
- *Analytical Geometry*
- *Data Augmentation*

Sinisa Colic and Samin Aref

Announcements

Mid-term assessment logistics

- Platform: Crowdmark
- Mock Assessment is available now (emailed to you) to get familiar with the logistics. It will not be graded. Due date Feb 10 at 9:00am.
- Some practice problems and solutions from past semesters are on Quercus

- Material in midterm:
  - Week 1 to 5 of Lectures, Tutorials 1 and 2, Project 1, Reading assignments 1-4.

Mid-term assessment logistics

- Midterm Assessment Distribution: **Tuesday, Feb 15 at 9am (Toronto time)**
- Deadline for Submitting the Assessment: **Wednesday, Feb 16 at 3pm (Toronto time)**
  - A countdown starts as soon as you access the assessment (**2.5 hours available** as per course schedule)
  - 30 minutes are already added to the countdown time for contingency. No excuses.
- Time needed for writing answers: about 90 minutes
- Late submission or no submission: 0 mark (as per syllabus).
- Use course material + online resources – NO HELP FROM OTHERS!! No Piazza.
- If needed, make assumptions and answer the questions. Do not contact us to ask.
- In case of a **logistic problem**, you can email Sinisa, Samin and Ali during your assessment. There is no guarantee that we can respond before your time runs out.

Slide Attribution

These slides contain materials from various sources. Special thanks to the following authors:

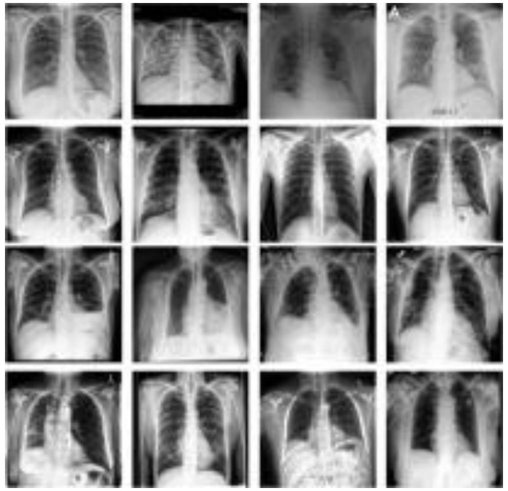
- Marc Deisenroth
- Mark Schmidt
- Jason Riordon

Last Time

- Looked into assessing model performance
  - Probability Theory
  - Gaussians
  - Confusion Matrix
  - ROCs

- Data Processing to the Rescue
  - Data Augmentation (today)
  - Dimensionality Reduction (next week)

What do these datasets have in common?



- How can we improve these datasets?

ML Performance Benchmarks

Method	Depth	Params	C10	C10+	C100	C100+
Network in Network [22]	-	-	10.41	8.81	35.68	-
All-CNN [32]	-	-	9.08	7.25	-	33.71
Deeply Supervised Net [20]	-	-	9.69	7.97	-	34.57
Highway Network [34]	-	-	-	7.72	-	32.39
FractalNet [17]	21	38.6M	10.18	5.22	35.34	23.30
with Dropout/Drop-path	21	38.6M	7.33	4.60	28.20	23.73
ResNet [11]	110	1.7M	-	6.61	-	-
ResNet (reported by [13])	110	1.7M	13.63	6.41	44.74	27.22
ResNet with Stochastic Depth [13]	110	1.7M	11.66	5.23	37.80	24.58
	1202	10.2M	-	4.91	-	-
Wide ResNet [42]	16	11.0M	-	4.81	-	22.07
	28	36.5M	-	4.17	-	20.50
with Dropout	16	2.7M	-	-	-	-
ResNet (pre-activation) [12]	164	1.7M	11.26*	5.46	35.58*	24.33
	1001	10.2M	10.56*	4.62	33.47*	22.71
DenseNet ($k = 12$)	40	1.0M	**7.00**	5.24	**27.55**	24.42
DenseNet ($k = 12$)	100	7.0M	**5.77**	**4.10**	**23.79**	**20.20**
DenseNet ($k = 24$)	100	27.2M	**5.83**	**3.74**	**23.42**	**19.25**
DenseNet-BC ($k = 12$)	100	0.8M	**5.92**	4.51	**24.15**	22.27
DenseNet-BC ($k = 24$)	250	15.3M	**5.19**	**3.62**	**19.64**	**17.60**
DenseNet-BC ($k = 40$)	190	25.6M	-	**3.46**	-	**17.18**

Error rates of popular neural networks on the Cifar 10 and Cifar 100 datasets. (Source: DenseNet)

- C10+ and C100+ highlight the error rates after data augmentation
- Data augmentation found to **consistently lower the error rates**!

Agenda

- Linear Algebra
  - Scalars, Vectors, Matrices
  - Solving Systems of Linear Equations
  - Linear Independence
  - Linear Mappings
- Analytic Geometry
  - Norms, Inner Products, Lengths, etc.
  - Angles and Orthonormal Basis
- Data Augmentation

Today's Theme:
Data Processing

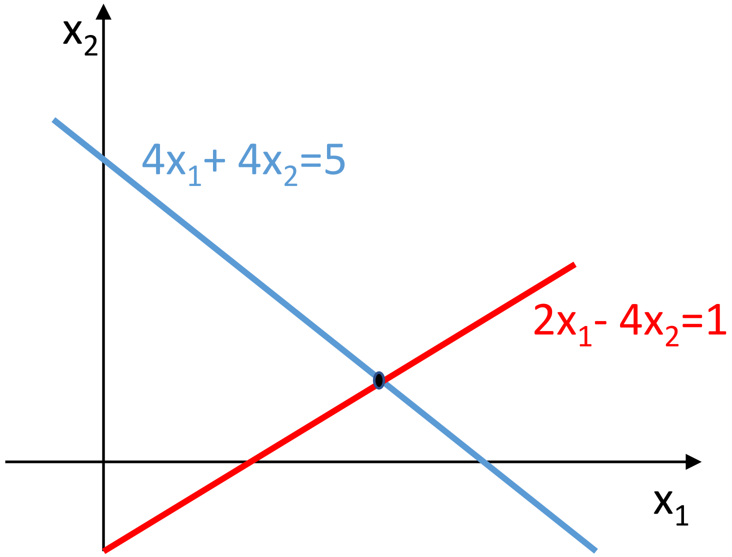
Linear Algebra

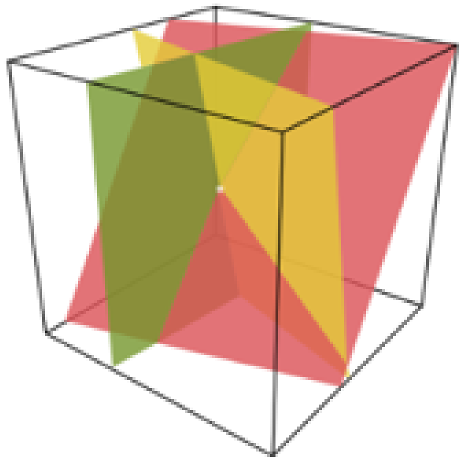
Readings:

- **Chapter 2.1-5 MML Textbook**

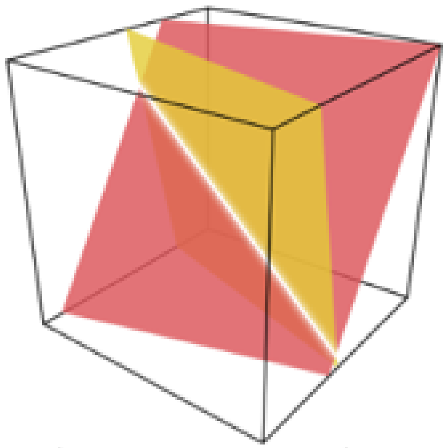
Systems of Linear Equations

- The solution space of a system of two linear equations with two variables can be geometrically interpreted as the intersection of two lines
- intersection of planes in three variables





System in three variables – solution is at intersection



System with 2 equations and three variables – solution is typically a line

Matrix Representation

- Used to solve systems of linear equations more systematically
- Compact notation collects coefficients into vectors, and vectors into matrices:

$$x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + \ \dots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Matrix Notation

- A matrix has m x n elements (with $m, n \in \mathbb{N}$, and a_{ij}, i=1,…,m; j=1,…,n) which are ordered according to a rectangular scheme consisting of m rows and n columns:

$$A = \begin{array}{cc} & \begin{array}{cccc} 1 & 2 & & n \end{array} \\ & \left[\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array}\right] \begin{array}{c} 1 \\ 2 \\ \\ m \end{array} \end{array} \quad a_{ij} \in \mathbb{R}$$

- By convention (1 by n)-matrices are called **rows** and (m by 1)-matrices are called **columns**. These special matrices are also called row/column vectors.
- (1 by 1)-matrices are referred to as scalars

Addition and Scalar Multiplication

- Vector addition:

$$a + b = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \end{bmatrix}$$

- Scalar multiplication:

$$\alpha b = \alpha \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \alpha b_1 \\ \alpha b_2 \end{bmatrix}$$

Addition and Scalar Multiplication

- Matrix addition: The **sum of two matrices** $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{m \times n}$ is defined as the element-wise sum:

$$A + B = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

- Scalar multiplication of **a matrix** $A \in \mathbb{R}^{m \times n}$ is defined as:

$$\alpha * A = \begin{bmatrix} \alpha * a_{11} & \dots & \alpha * a_{1n} \\ \vdots & & \vdots \\ \alpha * a_{m1} & \dots & \alpha * a_{mn} \end{bmatrix}$$

Matrix Multiplication

- We can multiply a matrix by a column vector:

$$Ax = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{bmatrix}$$

- We can multiply a matrix by a row vector:

$$x^T A = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{21}x_2 + a_{31}x_3 & a_{12}x_1 + a_{22}x_2 + a_{32}x_3 & a_{13}x_1 + a_{23}x_2 + a_{33}x_3 \end{bmatrix}$$

- In general, we can multiply matrices A and B when the number of columns in A matches the number of rows in B:

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

Example: Matrix Multiplication

- For two matrices: $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 3}, B = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2},$
- we obtain:

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \in \mathbb{R}^{2\times 2},$$

$$BA = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 2 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

Not commutative! $AB \neq BA$

Basic Properties

- A few properties:
  - **Associativity**:

$$\forall A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}, C \in \mathbb{R}^{p\times q}: (AB)C = A(BC)$$

  - **Distributivity**:

$$\forall A, B \in \mathbb{R}^{m\times n}, C, D \in \mathbb{R}^{n\times p}: \ (A + B)C = AC + BC$$

$$A(C + D) = AC + AD$$

Inner Product and Outer Product

- The **inner product** between vectors of the same length is:

$$a^T b = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \ \ldots + anbn = \gamma$$

The inner product is a **scalar**

- The **outer product** between vectors of the same length is:

$$ab^T = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \ldots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \ldots & a_2 b_n \\ \vdots & \vdots & & \vdots \\ a_n b_1 & a_n b_2 & \ldots & a_n b_n \end{bmatrix}$$

The outer product is a **matrix**

Identity Matrix

- We define the **identity matrix** as shown:

$$I_n := \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

- Any matrix multiplied by the identity will not change the matrix:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \text{X} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+0+0 & 0+2+0 & 0+0+3 \\ 4+0+0 & 0+5+0 & 0+0+6 \\ 7+0+0 & 0+8+0 & 0+0+9 \end{bmatrix}$$

Inverse

- If square matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ have the property that $AB = I_n = BA$. Then B is called the inverse of A and denoted by A^{-1}.

- Example, these matrices are inverse to each other:

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 4 & 4 & 5 \\ 6 & 7 & 7 \end{bmatrix}, \qquad B = \begin{bmatrix} -7 & -7 & 6 \\ 2 & 1 & -1 \\ 4 & 5 & -4 \end{bmatrix}$$

- **We'll look at how to calculate the inverse later**

Transpose

- **Transpose definition**: For $A \in \mathbb{R}^{m\times n}$ the matrix $B \in \mathbb{R}^{n\times m}$ with $b_{ij} = a_{ji}$ is called transpose of A. We write $B = A^T$.
- **Symmetric Matrix**: A matrix $A \in \mathbb{R}^{n\times n}$ is symmetric if $A = A^T$.
- Some useful identities:

$$AA^{-1} = I = A^{-1}A$$
$$(AB)^{-1} = B^{-1}A^{-1}$$
$$(A+B)^{-1} \neq A^{-1} + B^{-1}$$
$$(A^T)^T = A$$
$$(A+B)^T = A^T + B^T$$
$$(AB)^T = B^T A^T$$

Solving Systems of Linear Equations

- Given A and b, we want to solve for x:

$$Ax = b \qquad \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}$$

- Key to solving a system of linear equations are elementary transformations that keep the solution set the same, but that **transform the equation system into a simpler form**.
    1. Exchange of two equations (rows in the matrix)
    2. Multiplication of an equation (row) with a constant
    3. Addition of two equations (rows)

- This leads us to Gaussian Elimination (aka row reduction)

Triangular Linear Systems

- Consider a square linear system with an **upper triangular matrix** (non-zero diagonals):

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

- We can solve this system bottom to top using substitution:

$$a_{33}x_3 = b_3 \qquad x_3 = \frac{b_3}{a_{33}}$$

$$a_{22}x_2 + a_{23}x_3 = b_2 \qquad x_2 = \frac{b_2 - a_{23}x_3}{a_{22}}$$

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \qquad x_1 = \frac{b_1 - a_{13}x_3 - a_{12}x_2}{a_{11}}$$

Example: Gaussian Elimination

- Gaussian elimination uses elementary row operations to transform a linear system into a triangular system:

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 5 \\ 4x_1 - 6x_2 &= -2 \\ -2x_1 + 7x_2 + 2x_3 &= 9 \end{aligned} \longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right]$$

  - Add <u>-2 times first row</u> to second
  - Add <u>1 times first row</u> to third

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 5 \\ -8x_2 - 2x_3 &= -12 \\ 8x_2 + 3x_3 &= 14 \end{aligned} \longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 8 & 3 & 14 \end{array}\right]$$

  - Add <u>1 times second row</u> to third

$$\boxed{\begin{aligned} 2x_1 + x_2 + x_3 &= 5 \\ -8x_2 - 2x_3 &= -12 \\ x_3 &= 2 \end{aligned}} \longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

Row Echelon form

Row Echelon Form

- The first non-zero coefficient from the left (the "leading coefficient") is always to the right of the first non-zero coefficient in the row above.
- Rows consisting of all zero coefficients are at the bottom of the matrix.

$$\left[\begin{array}{ccc|c} \boxed{2} & 1 & 1 & 5 \\ 0 & \boxed{-8} & -2 & -12 \\ 0 & 0 & \boxed{1} & 2 \end{array}\right]$$

Row Echelon form

Example: Reduced Echelon Form

- We can simplify this even further:

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 5 \\ -8x_2 - 2x_3 &= -12 \\ x_3 &= 2 \end{aligned} \longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

  - Divide first row by 2
  - Divide 2nd row by -8

$$\begin{aligned} x_1 + 0.5x_2 + 0.5x_3 &= 2.5 \\ x_2 + 0.25x_3 &= 1.5 \\ x_3 &= 2 \end{aligned} \longrightarrow \left[\begin{array}{ccc|c} 1 & 0.5 & 0.5 & 2.5 \\ 0 & 1 & 0.25 & 1.5 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

  - Add <u>-0.25 times third row</u> to second row
  - Add <u>-0.5 times third row</u> to first row
  - Add <u>-0.5 times second row</u> to first row

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 1 \\ x_3 &= 2 \end{aligned} \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

Reduced row Echelon form

The coefficient matrix could be non-square Example 2.6 from MML book (reading assignment 4):

For $a \in \mathbb{R}$, we seek $\boxed{\text{all solutions}}$ of the following system of equations:

$$\begin{array}{rcrcrcrcrcr} -2x_1 & + & 4x_2 & - & 2x_3 & - & x_4 & + & 4x_5 & = & -3 \\ 4x_1 & - & 8x_2 & + & 3x_3 & - & 3x_4 & + & x_5 & = & 2 \\ x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 \\ x_1 & - & 2x_2 & & & - & 3x_4 & + & 4x_5 & = & a \end{array} . \tag{2.44}$$

Four equations and five unknowns

Alternative Method: Inverse Matrix

- We can also solve linear systems of equations by applying the inverse.
- The solution to $Ax = b$ can be obtained by multiplying by A^{-1} to isolate for x.

$$
\begin{aligned}
Ax &= b \\
A^{-1}Ax &= A^{-1}b \\
I_n x &= A^{-1}b \\
x &= A^{-1}b
\end{aligned}
$$

Note that A^{-1}will cancel out A only if multiplied from the left-hand side, otherwise we have $A^{-1}xA$

Calculating an Inverse Matrix

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

- Write down the augmented matrix with the identity on the right-hand side

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

- **Apply Gaussian elimination** to bring it into **reduced row-echelon form**.

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -1 & 2 & -2 & 2 \\ 0 & 1 & 0 & 0 & 1 & -1 & 2 & -2 \\ 0 & 0 & 1 & 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 2 \end{array}\right]$$

The desired inverse is given as its right-hand side.

We can verify that this is indeed the inverse by performing the multiplication AA^{-1} and observing that we recover I_n.

$$A^{-1} = \begin{bmatrix} -1 & 2 & -2 & 2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 1 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$

What can go wrong?

- Applying Gaussian Elimination (row reduction) does not always lead to a solution.
- **Singular Case:** When we have a 0 in a pivot column. This is an example of a matrix that is not invertible.
- For example:

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & \boxed{0} & 3 & -2 \\ 0 & 0 & 4 & 2 \end{array}\right] \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & \boxed{0} & 0 \end{array}\right]$$

singular   singular

- To understand this better it helps to consider matrices from a geometric perspective.

Several Interpretations

- Given A and b, we want to solve for x:

$$Ax = b \qquad \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

- This can be given several interpretations:
  - **By rows**: x is the intersection of hyper-planes:

$$2x - y = 1$$
$$x + y = 5$$

  - **By columns**: x is the linear combination that gives b:

$$x \begin{bmatrix} 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

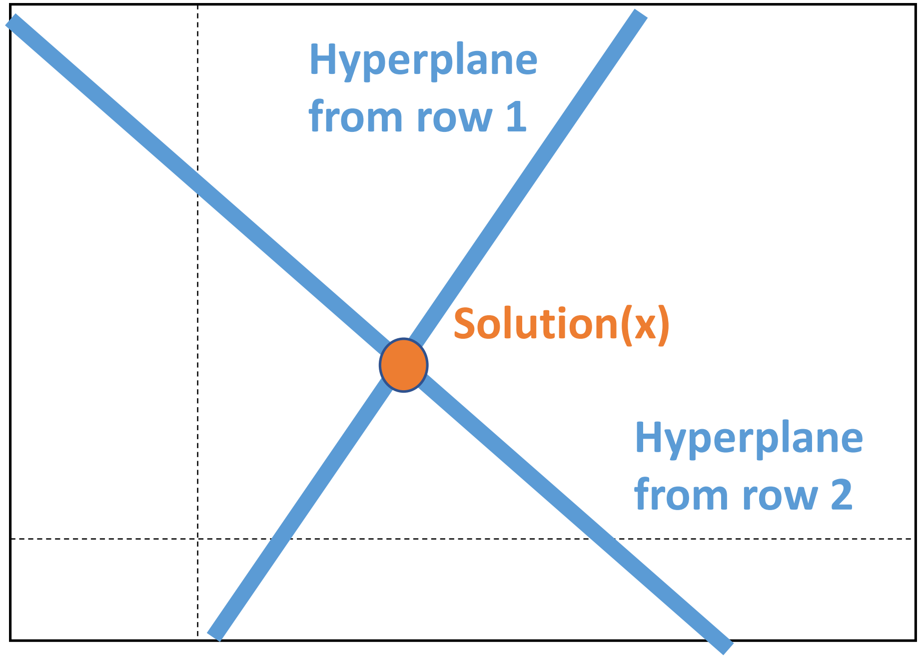
  - **Transformation**: x is the vector transformed to b:

$$T(x) = b$$

Geometry of Linear Equations

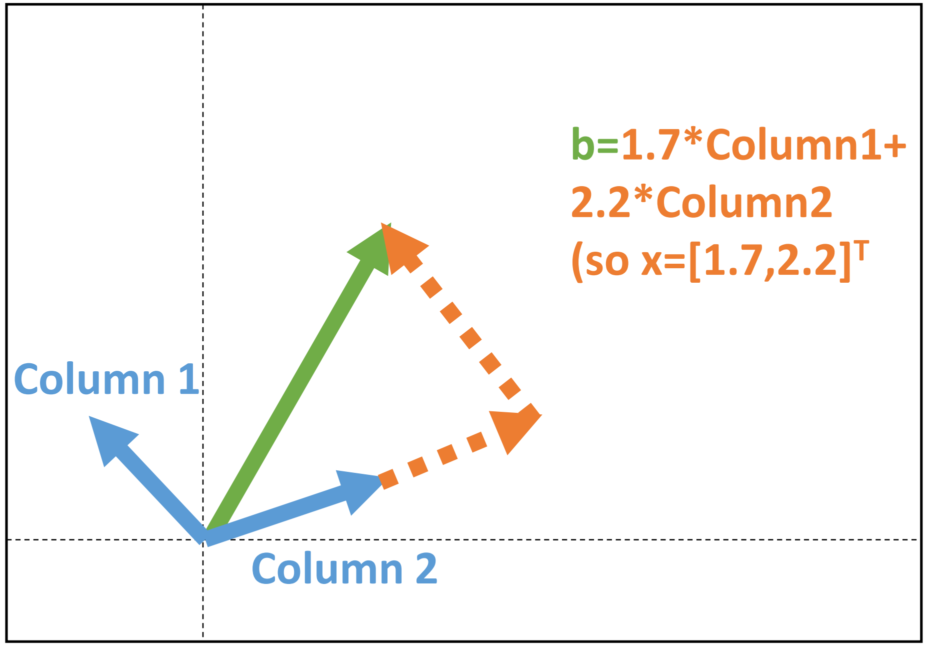
➢ By Rows:

Find intersection of hyperplanes



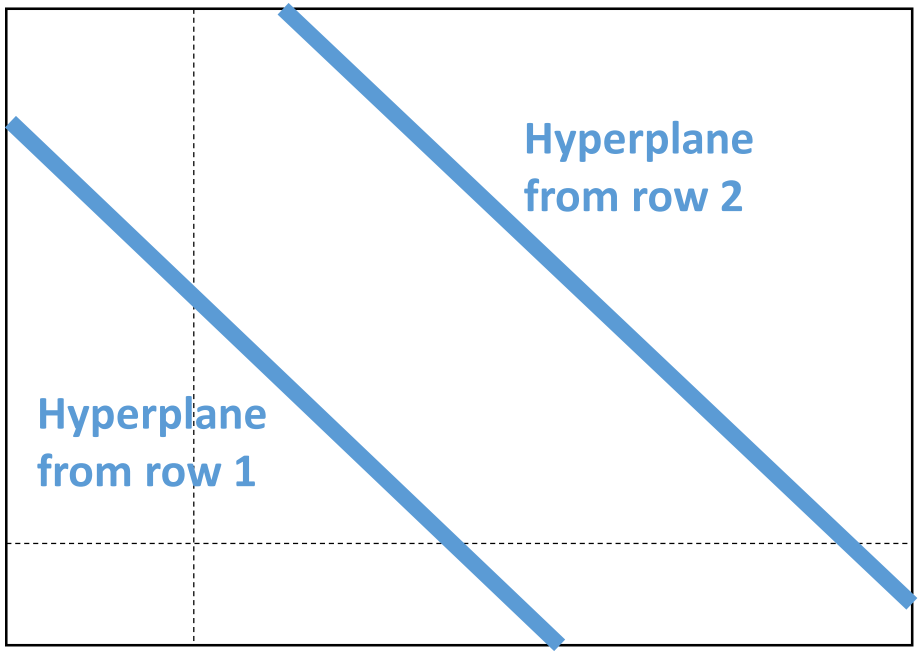
➢ By Columns:

Find linear combination of columns

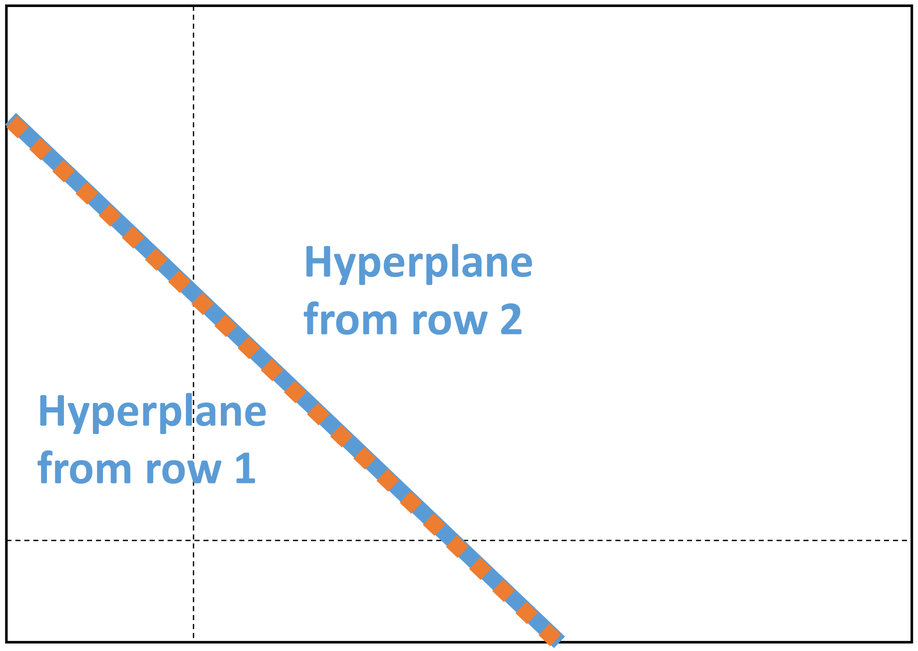


What can go wrong?

- By rows:



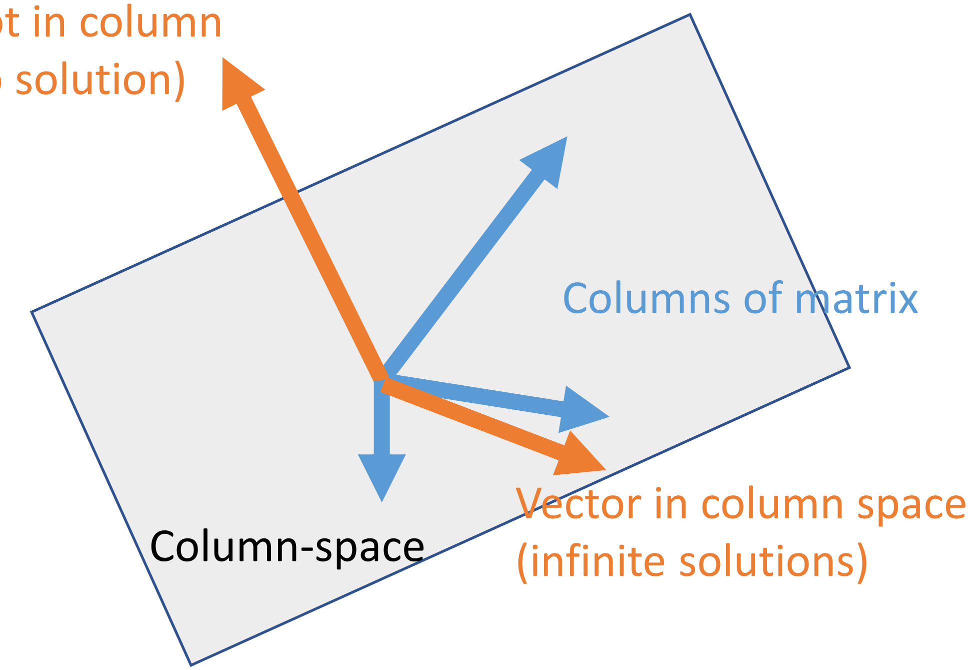
No intersection



Infinite intersection

What can go wrong?

- By columns:

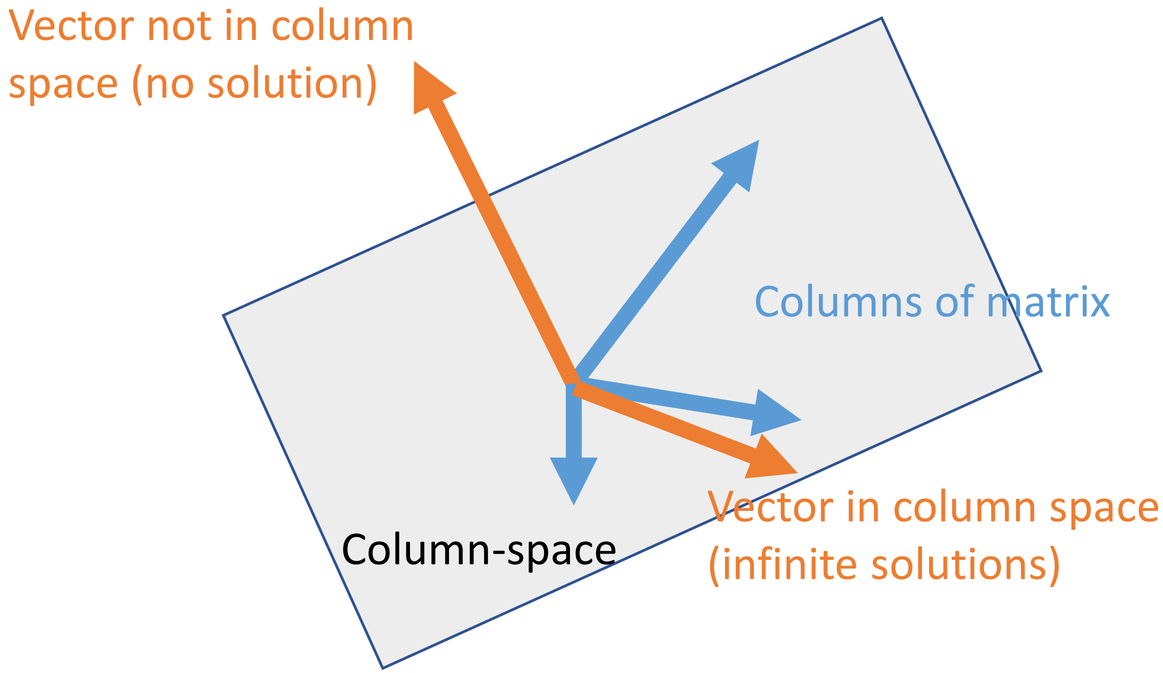


Solutions to Ax=b

- Q: In general, when does Ax=b have a unique solution?

- A: When b is in the column-space of A, and the columns of A are **linearly independent**

- Q: What does it mean to be independent?



Linear Dependence

- A set of vectors is either linearly dependent or linearly independent.
- A vector is linearly dependent on a set of vectors if it can be written as a linear combination of them:

$$c = \alpha_1 b_1 + \alpha_2 b_2 + \ \dots \alpha_n b_n$$

- We say that c is "linearly dependent" on $\{b_1, b_2, \dots, b_3\}$, and that the set $\{c, b_1, b_2, \dots, b_3\}$ is "linearly dependent"
- A set is linearly dependent iff the zero vector can be written as a combination of the vectors $\{b_1, b_2, \dots, b_3\}$:

$$\exists_\alpha \neq 0, s.t. 0 = \alpha_1 b_1 + \alpha_2 b_2 + \ \dots \alpha_n b_n \Rightarrow \{b_1, b_2, \dots, b_n\}\ dependent$$

Linear Independence

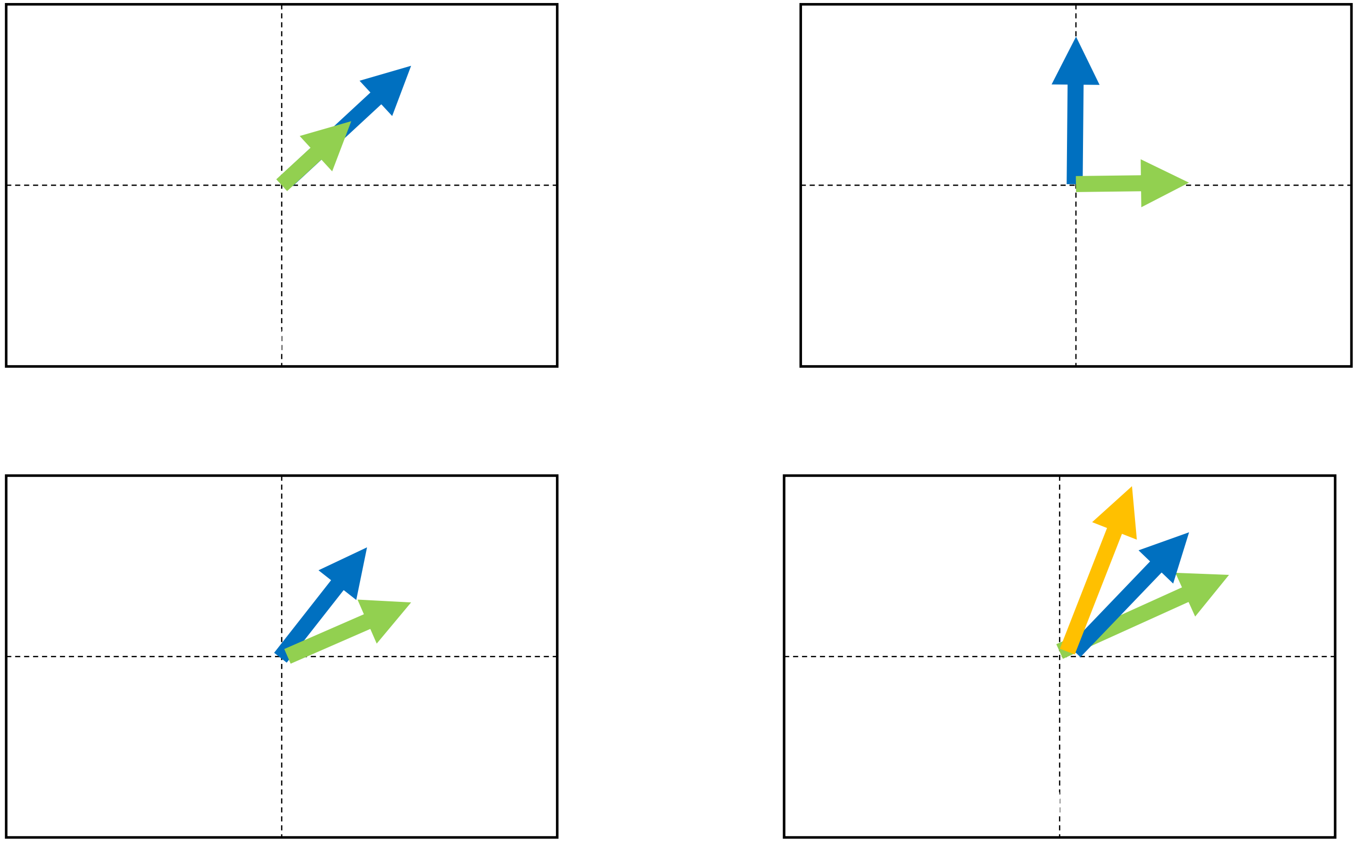
- If a set of vectors is not linearly dependent, we say it is linearly independent
- The zero vector **cannot** be written as a combination of independent vectors unless all coefficients α are set to zero:

$$0 = \alpha_1 b_1 + \alpha_2 b_2 + \ \dots \alpha_n b_n => \ \alpha_i = 0 \ \forall_i$$

- If the **vectors are independent**, then **there is no way represent one of the vectors as a combination of the others**.

Linear Dependence vs Independence

- Q: Determine independence in R^2 for the following.



Linear Independence

- Consider we have a set of three vectors $\{x_1, x_2\,, x_3\} \,\epsilon\, \mathbb{R}^4$
- To check whether they are linearly dependent, we solve: $\lambda_1 x_1 + \lambda_2 x_2 + \lambda_2 x_3 = 0$
- We write the vectors $x_i, i = 1, 2, 3$, as the columns of a matrix and **apply elementary row operations** until we identify the pivot columns.
- All column vectors are **linearly independent if and only if all columns are pivot columns.** If there is at least one non-pivot column the vectors are linearly dependent.

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}, x_1 = \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix}$$

$$\lambda_1 x_1 + \lambda_2 x_2 + \lambda_2 x_3 = 0$$

$$\begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & -2 \\ -3 & 0 & 1 \\ 4 & 2 & 1 \end{bmatrix}$$

$$\cdots$$

$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Vector Space

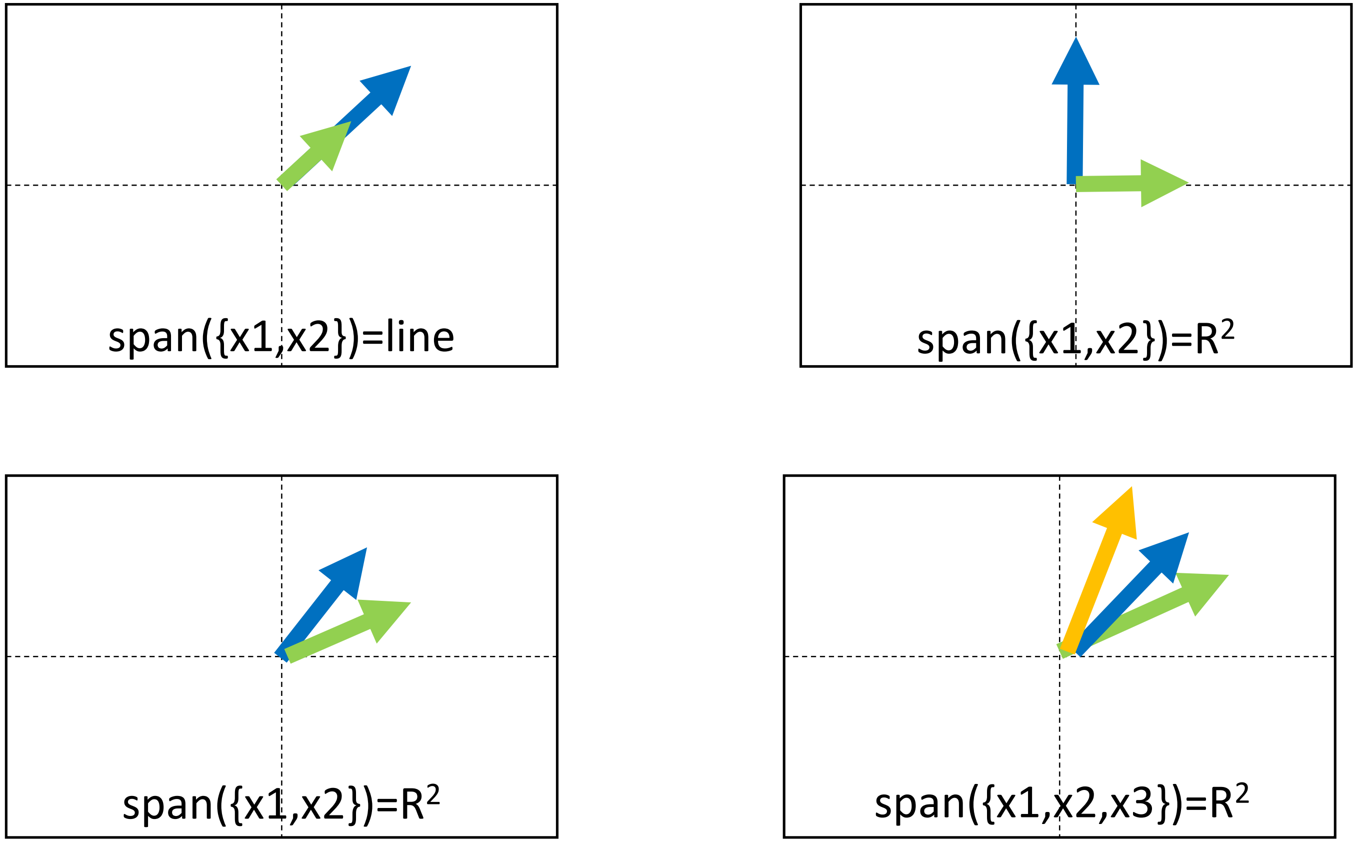
- A vector space is a set of objects called “vectors”, with closed operations “addition” and “scalar multiplication” satisfying certain axioms:

  1. $x + y = y + x$
  2. $x + (y + z) = (x + y) + z$
  3. *exists a zero vector “0” s.t.* $\forall_x, x + 0 = x$
  4. $\forall_x, exists\ an\ additive\ inverse\ \text{“}-x\text{”}, s.t.\ \ x + (-x) = 0$
  5. $1x = x$
  6. $(c_1 c_2)x = c_1(c_2 x)$
  7. $c(x + y) = cx + cy$
  8. $(c_1 + c_2)x = c_1 x + c_2 x$

- Examples: $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^n$

Subspace

- Subspaces generated in R^2:



set of vectors $\mathcal{A} = \{x_1, \ldots, x_k\} \subseteq \mathcal{V}$

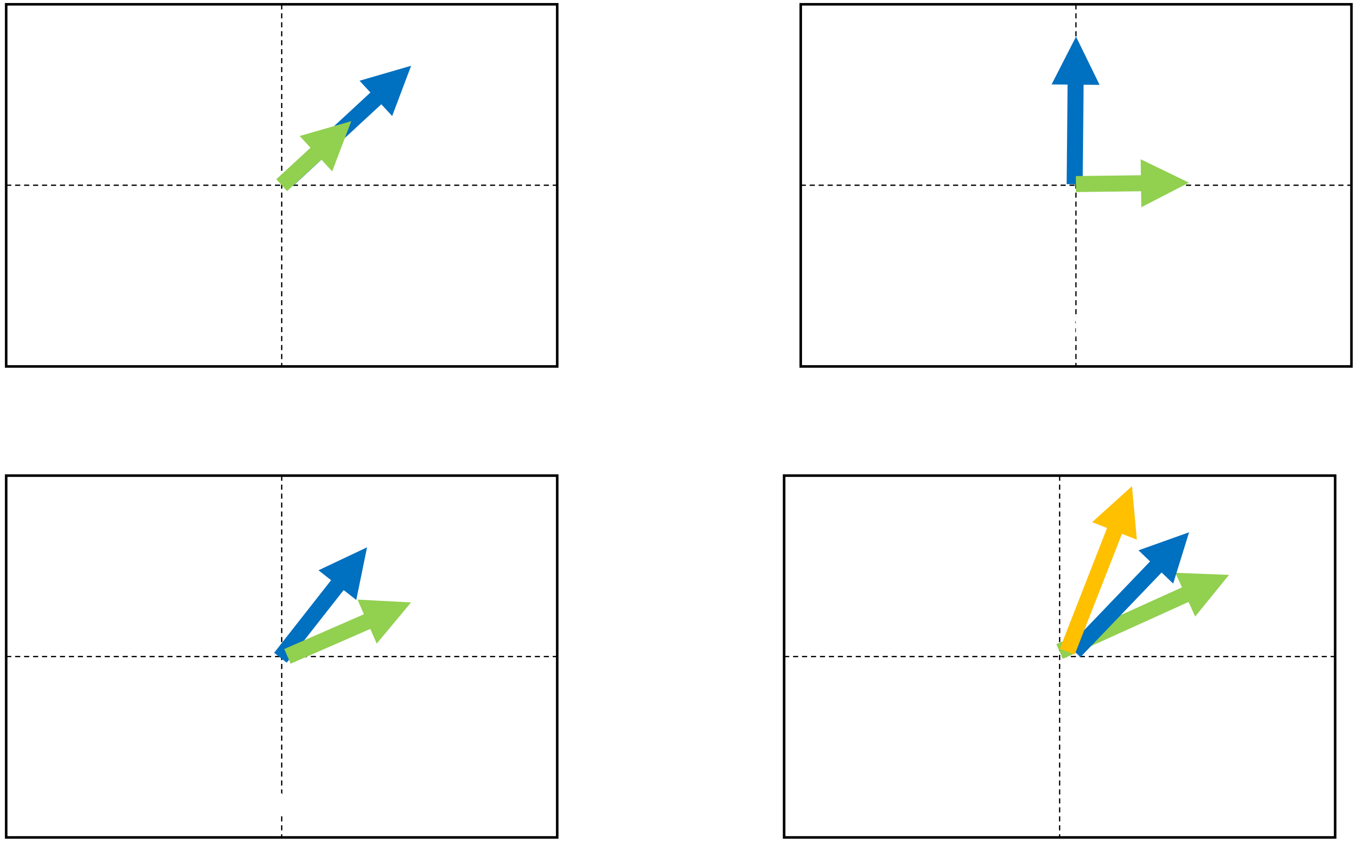
The set of all linear combinations of vectors in $\mathcal{A}$ is called the span of $\mathcal{A}$. If $\mathcal{A}$ spans the vector space V, write $V = span[\mathcal{A}]$ or $V = span[x_1, \ldots, x_k]$

Basis

- The vectors that span a subspace are not unique
- However, the minimum number of vectors needed to span a subspace is unique
- This number is called the **dimension or rank of the subspace**
- A **minimal set of vectors that span a subspace is called a basis** for the space
- The **vectors in a basis must be linearly independent**, otherwise we could remove one and still span space

Basis

- Basis in vector space $V \in R^2$:



Every **linearly independent set of vectors** that span V is called a basis of V

Example Bases

- In $\mathbb{R}^3$, the **canonical/standard basis** is:

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

- Two different bases of $\mathbb{R}^3$are:

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} \quad \mathcal{B}_2 = \left\{ \begin{bmatrix} 0.5 \\ 0.8 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 1.8 \\ 0.3 \\ 0.3 \end{bmatrix}, \begin{bmatrix} -2.2 \\ -1.3 \\ 3.5 \end{bmatrix} \right\}$$

Linear Mapping/Transformation

- Earlier, we saw that vectors are objects that can be added together and multiplied by a scalar, and the resulting object is still a vector
- Now, we do the same for vector spaces

- **Linear Mapping**: For vector spaces V, W, a mapping $\phi: V \rightarrow W$ is called a linear mapping (or linear transformation) if:

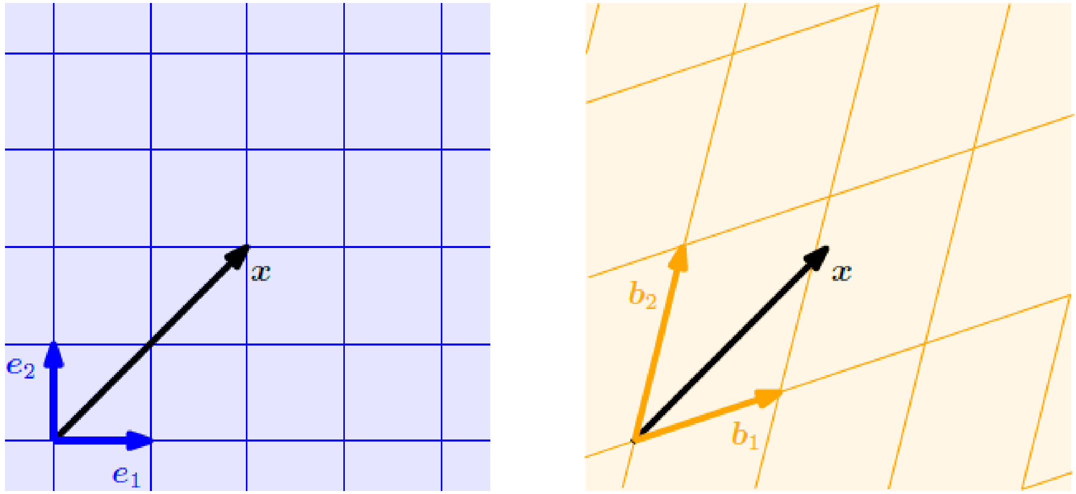
$$\forall x, y \in V \;\; \forall \lambda, \psi \in \mathbb{R}: \; \Phi(\lambda x + \psi y) = \lambda \Phi(x) + \psi \Phi(y)$$

It turns out that we can represent linear mappings as matrices. Recall that we can also collect a set of vectors as columns of a matrix. When working with matrices, we have to keep in mind what the matrix represents: **a linear mapping or a collection of vectors**.

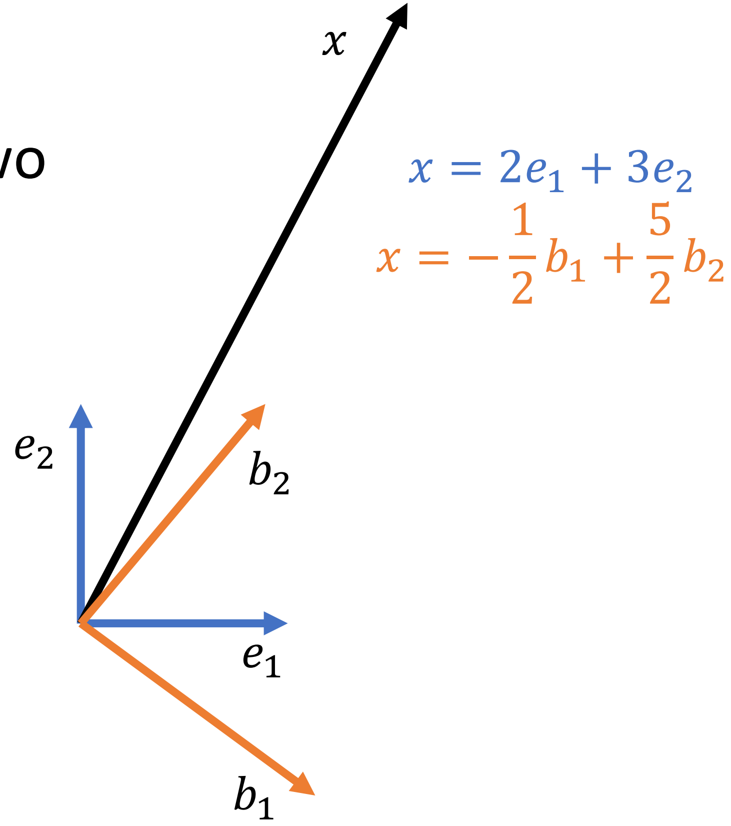
Linear Mapping/Transformation

- A vector has different coordinate representations depending on which coordinate system or basis is chosen.

- Example: two different coordinate systems defined by two sets of basis vectors.

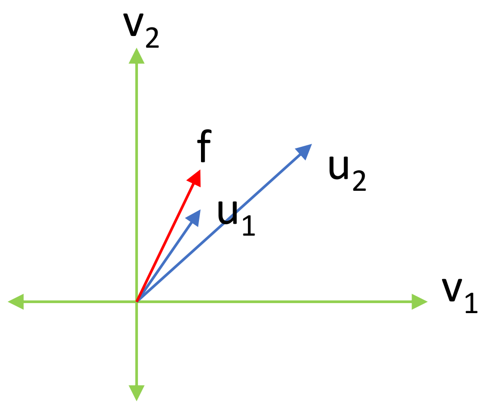


two different bases



Source: Eli Bendersky

Example: Change of Basis Matrix



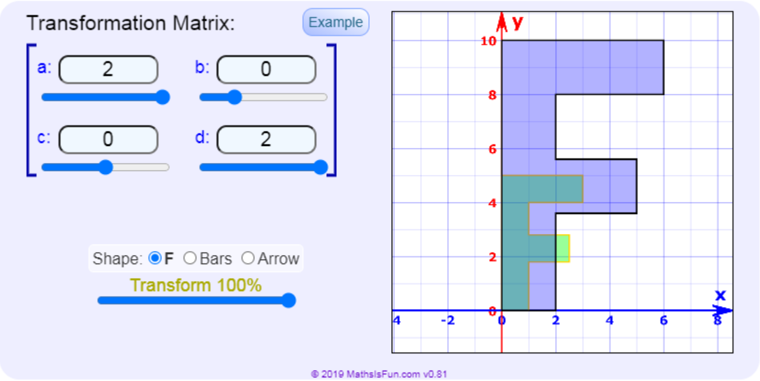
$U = [2\ 3]^T[4\ 5]^T$

$[f]_v=[2\ 4]^T$

$[f]_u=?$

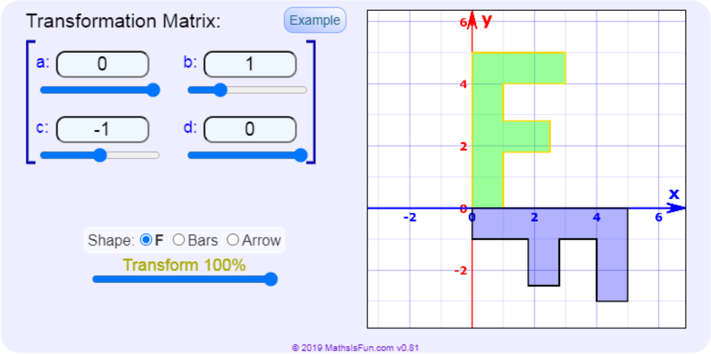
Examples of Transforms

- Scale

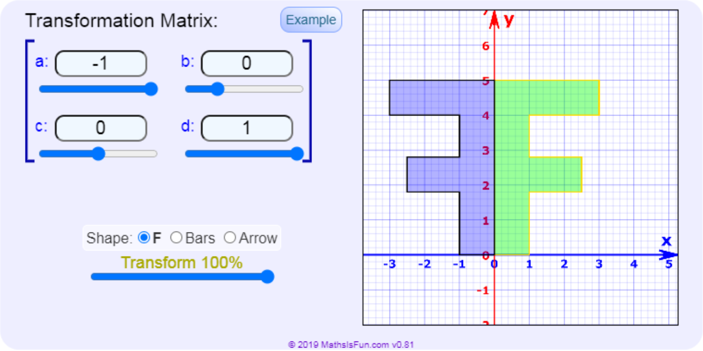


- Rotation

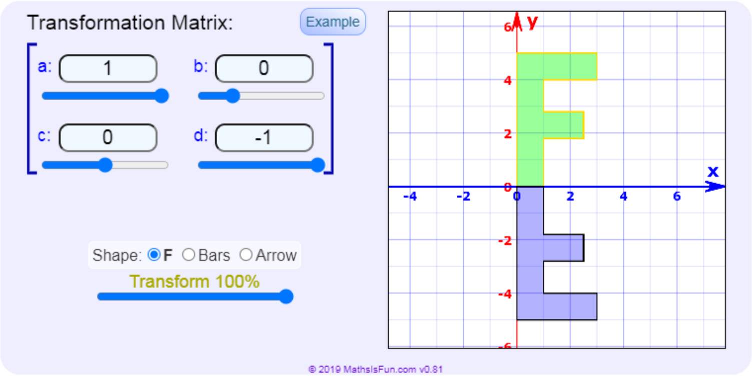
Have a play with this 2D transformation app:



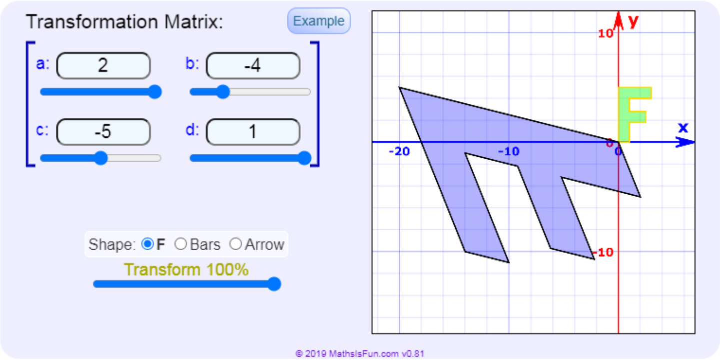
- Horizontal Mirror



- Vertical Mirror



- Combination of Transformations



Source: mathisfun.com

Short Break

Analytical Geometry

Readings:

- **Chapter 3.1-5,8,9 MML Textbook**

Norms

- A norm is a scalar measure of a vector's length.
- The most important norm is the Euclidean norm and for $x \in \mathbb{R}^n$ is defined as:

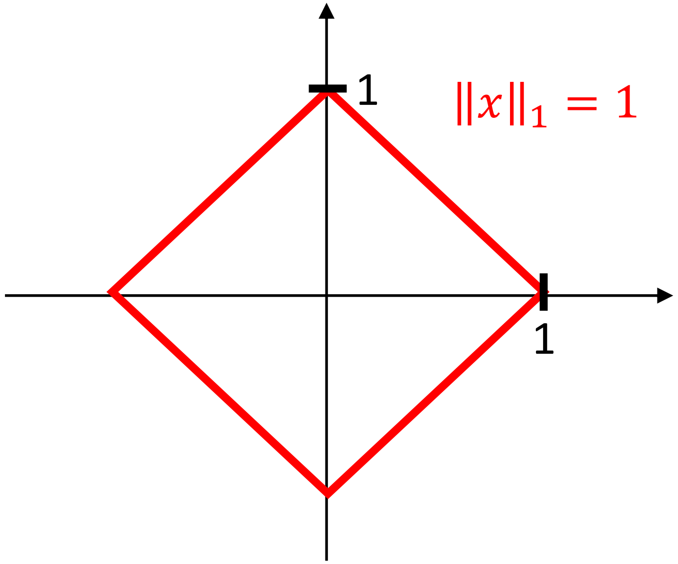
$$\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^T x}$$

computes the Euclidian distance of x from the origin.

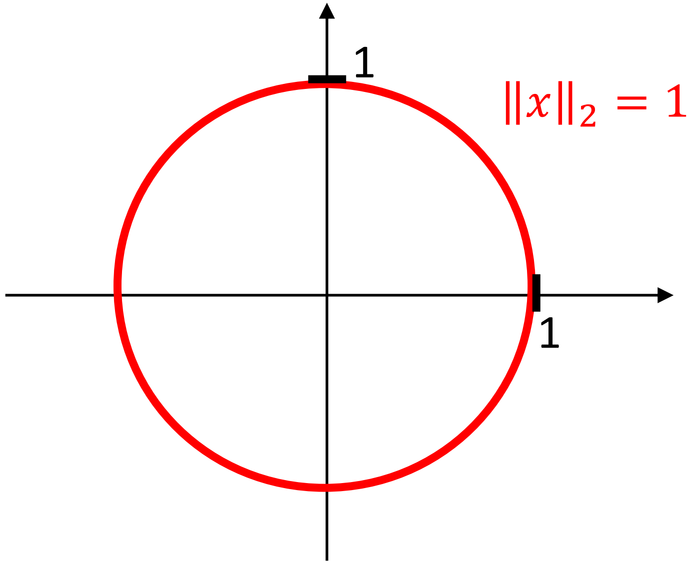
Euclidean norm is also known as the L2 norm

Norms

- For different norms, the red lines indicate the set of vectors with norm 1.



Manhattan norm



Euclidean distance

Dot product

- Dot product:

$$x^T y = \sum_{i=1}^{n} x_i y_i$$

$$a_1 \cdot b_1 = \begin{bmatrix} 1 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 5 \end{bmatrix} = (1 \cdot 3) + (7 \cdot 5) = 38$$

- Commonly, the dot product between two vectors a, b is denoted by $a^T b$ or $\langle a, b \rangle$.

Lengths and Distances

- Consider an inner product space.

- Then

$$d(x, y) := \|x - y\| = \sqrt{\langle x - y, x - y \rangle}$$

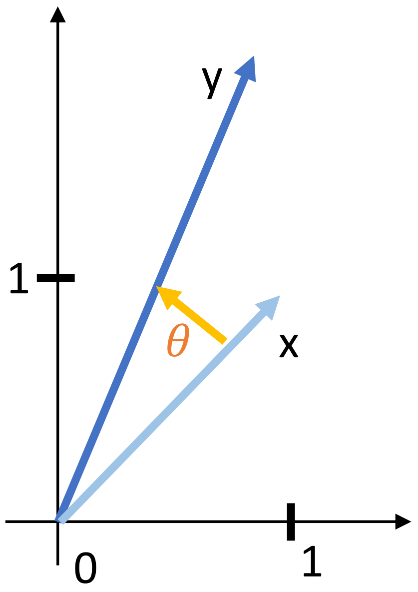
is called the distance between x and y for $x, y \in V$.

- If we use the dot product as the inner product, then the distance is called Euclidean distance.

Angles

- **The angle $\boldsymbol{\theta}$ between two vectors $\boldsymbol{x}, \boldsymbol{y}$ is computed using the inner product.**

- For Example: Let us compute the angle between $x = [1,1]^T \in \mathbb{R}^2$ and $y = [1,2]^T \in \mathbb{R}^2$



- Using the dot product as the inner product we get:

$$\cos\theta = \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle\langle y, y\rangle}} = \frac{x^T y}{\sqrt{x^T x y^T y}} = \frac{3}{\sqrt{10}}$$

- Then the angle between the two vectors is $\cos^{-1}(\frac{3}{\sqrt{10}}) \approx 0.32rad$, which corresponds to approximately 18°.

Orthogonality

- **Orthonormal = Orthogonal and unit vectors**

- Orthogonal Matrix: A square matrix $A \in \mathbb{R}^{n \times n}$ is an orthogonal matrix if and only if its columns are orthonormal so that

$$AA^T = I = A^T A,$$

- which implies that

$$A^{-1} = A^T,$$

i.e., the inverse is obtained by simply transposing the matrix.

Orthonormal Basis

- In n-dimensional space, we need n basis vectors that are linearly independent, if these vectors are orthogonal, and each has length 1, it’s a special case: **orthonormal basis**

- Consider an n-dimensional vector space V and a basis $\{b_1, \dots, b_n\}$ of V. If

$$\langle b_i, b_j \rangle = 0 \ for \ i \neq j$$
$$\langle b_i, b_i \rangle = 1$$

for all $i, j = 1, \dots, n$ then the basis is called an orthonormal basis (ONB). Note that $\langle b_i, b_i \rangle = 1$ implies that every basis vector has length/norm 1.

- If only $\langle b_i, b_j \rangle = 0 \ for \ i \neq j$ is satisfied, then the basis is called an orthogonal basis.

Orthonormal Basis

- The **canonical/standard basis for a Euclidean vector space $\mathbb{R}^n$ is an orthonormal basis**, where the inner product is the dot product of vectors.

- Example: In $\mathbb{R}^2$, the vectors:

$$b_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad b_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

form an orthonormal basis since $b_1^T b_2 = 0$ and $||b_1|| = 1 = ||b_2||$.

Orthogonal Projections

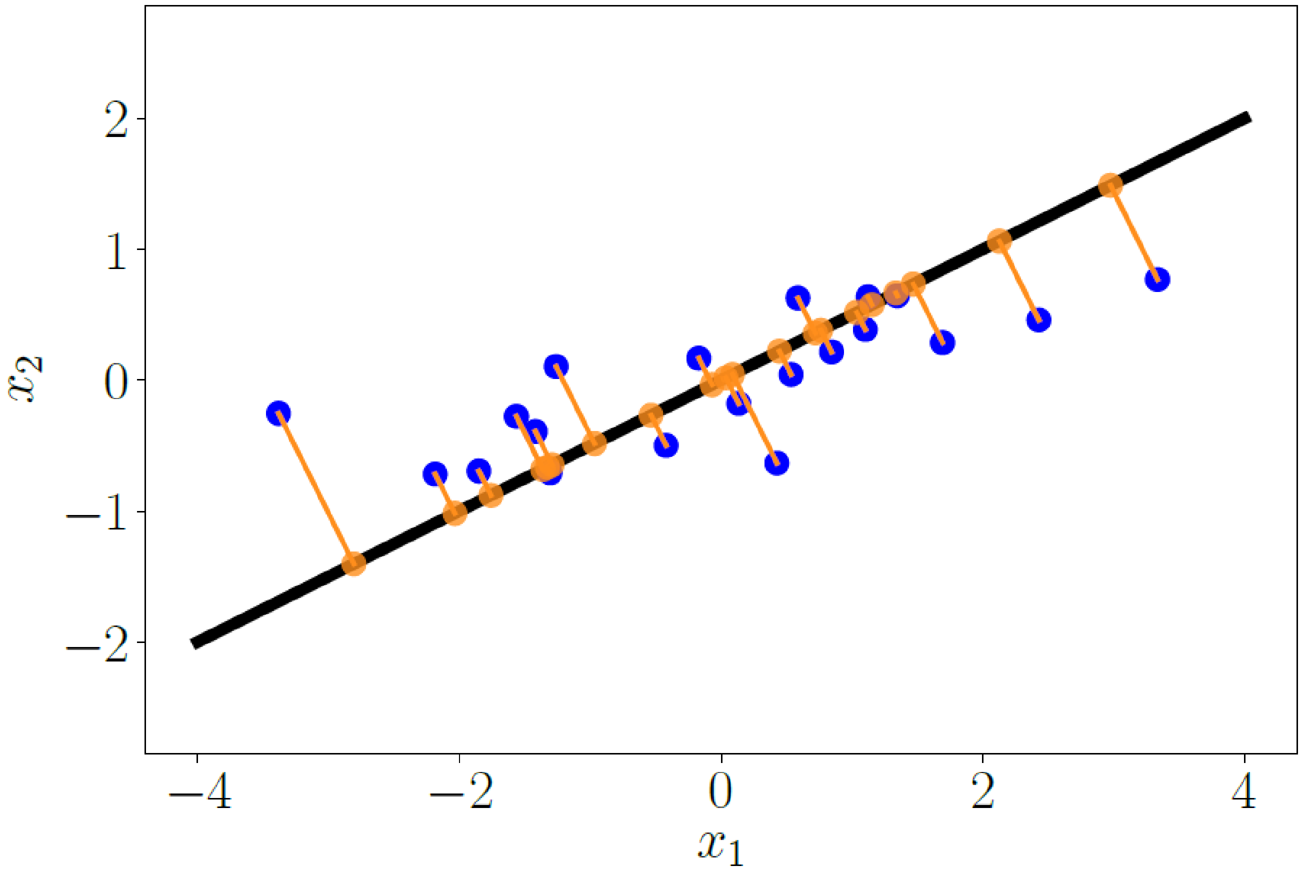
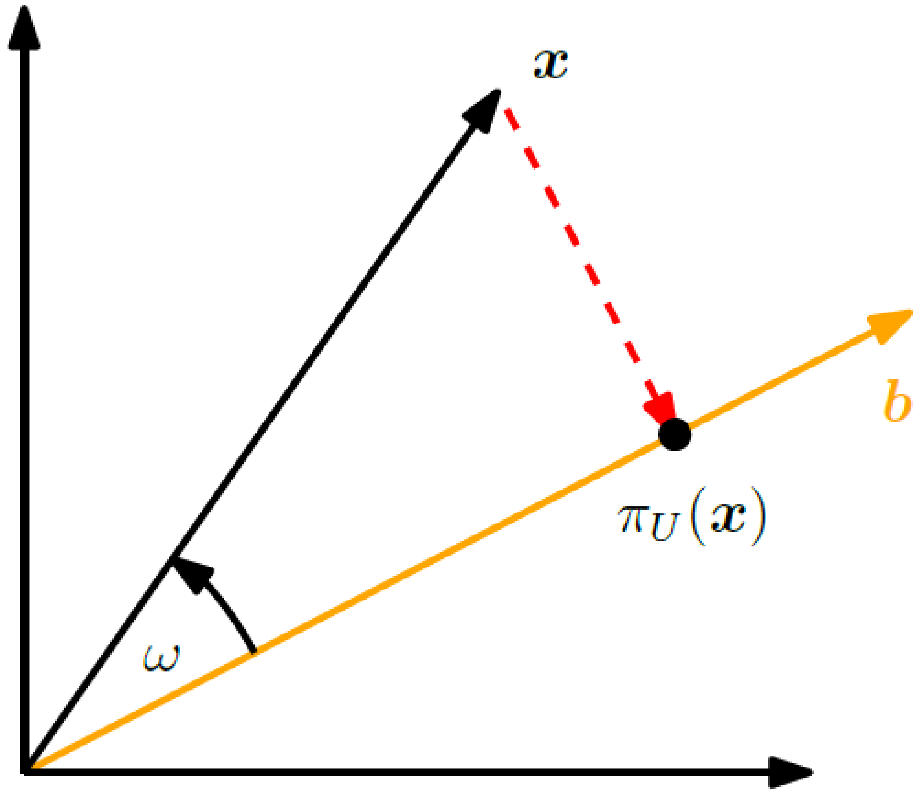


Figure 3.9 Orthogonal projection (orange dots) of a two-dimensional dataset (blue dots) onto a one-dimensional subspace (straight line).

- Projections are linear transformations, project to lower dimensional feature space

Orthogonal Projections

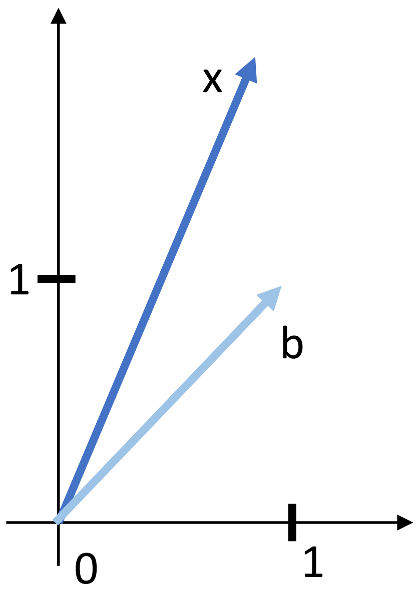


(a) Projection of $\boldsymbol{x} \in \mathbb{R}^2$ onto a subspace U with basis vector $\boldsymbol{b}$.

- The projection is defined

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \boldsymbol{b}\frac{\boldsymbol{b}^T\boldsymbol{x}}{\|\boldsymbol{b}\|^2} = \frac{\boldsymbol{b}\boldsymbol{b}^T}{\|\boldsymbol{b}\|^2}\boldsymbol{x}$$

Example: Orthogonal Projections



- Compute the projection of $x = [1,2]^T \in \mathbb{R}^2$ onto $b = [1,1]^T \in \mathbb{R}^2$

$$\pi_U(\boldsymbol{x}) = \frac{\boldsymbol{b}\boldsymbol{b}^T}{\|\boldsymbol{b}\|^2}\boldsymbol{x}$$

Projection Matrix

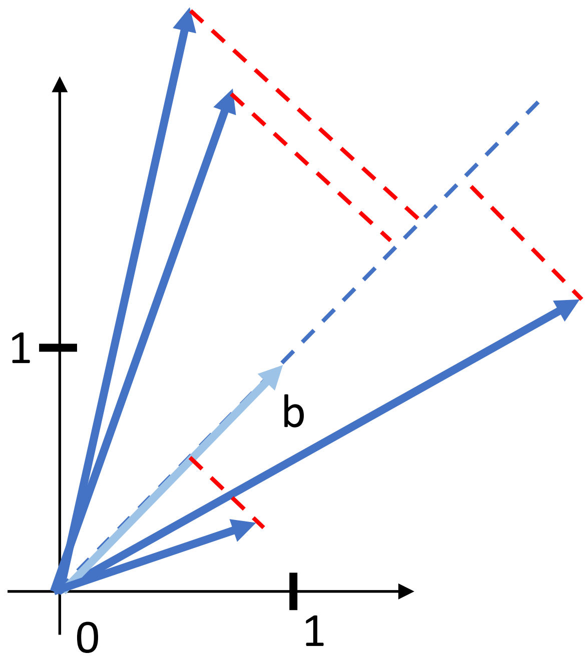
- We can also use a projection matrix, which allows us to project any vector x onto the subspace defined by π.

$$\pi_U(\boldsymbol{x}) = \frac{\boldsymbol{b}\boldsymbol{b}^T}{||\boldsymbol{b}||^2}\boldsymbol{x}$$

$$P_\pi = \boxed{\frac{\boldsymbol{b}\boldsymbol{b}^T}{||\boldsymbol{b}||^2}}$$

- Note that $\boldsymbol{b}\boldsymbol{b}^T$ will be a symmetric matrix

Example: Applying Projection Matrix



- Compute the projection matrix for $b = [1,1]^T \in \mathbb{R}^2$

$$P_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^T}{||\boldsymbol{b}||^2}$$

Examples in Google Colab

Data Augmentation

Non-Representative Data

- Everything our algorithms learn comes form the data used to train them.

- If the data is of **poor quality, unbalanced** or **not representative** of the task we want to solve, then how are our algorithms going to learn to generalize?

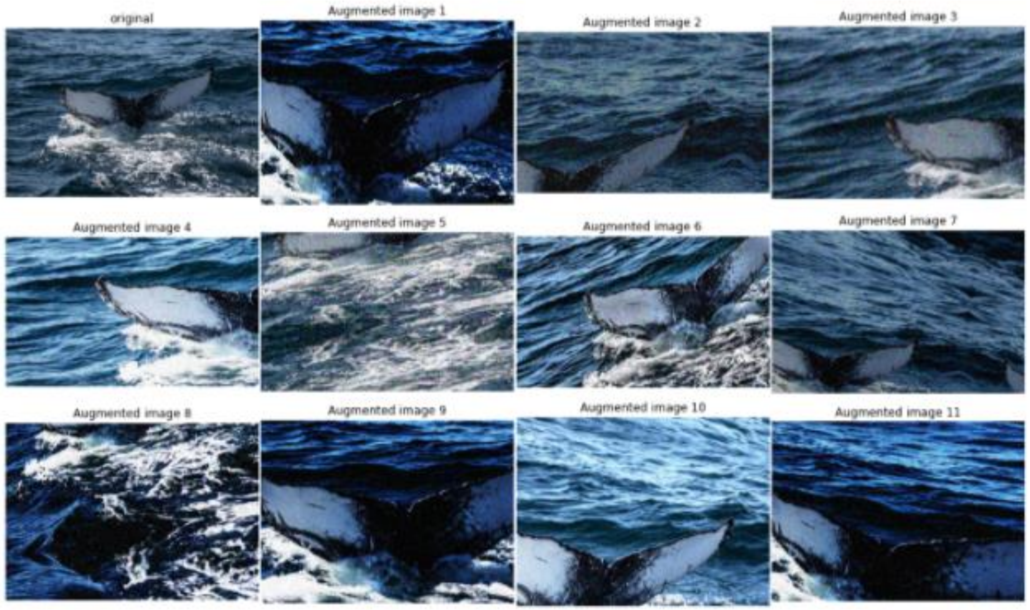
Capacity and Training

- Deep learning algorithms **have the capacity to classify real images** in various orientations and scales.

- If you train your algorithms on perfectly processed samples, then they won't know how to predict anything but perfectly cropped images.

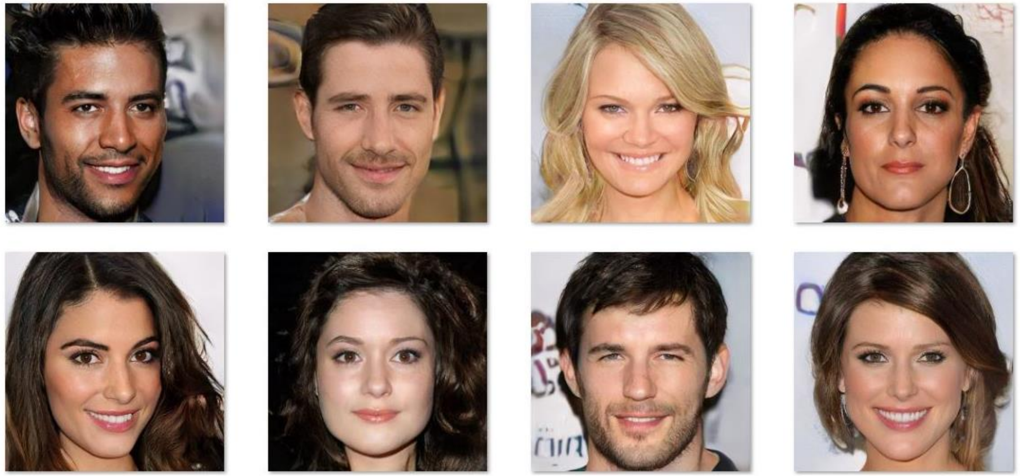
Data Augmentation

- Use linear algebra to perform common transformations to supplement datasets
  - Translation, Scaling, Rotation, Reflection
  - Noise, Light and Colour Intensity
  - Many more…



Source: kaggle.com

GAN Fake Celebrities



Source: Viridian Martinez

- Advanced:
  - Generative models (i.e., Deep learning) to create new images with similar characteristics

Test Time Data Augmentation

- You can also apply data augmentation to better evaluate your performance on test examples.
- Great way to assess limitations of your model to images of different rotations, scales, noise, etc.

Data Augmentation in Google Colab

Next Time

- Week 5 Tutorial 2 on Anomaly Detection
  - Project 2 Overview
- **Feb 15-16 Midterm (no lecture)**
- **Q&A Session on Thur. Feb 17**
- Project 2 is due on Feb 28th
- Week 7 – Dimensionality Reduction
  - Curse of Dimensionality
  - Eigendecomposition
  - Singular Value Decomposition
  - Principle Component Analysis