DSC 190

Machine Learning: Representations

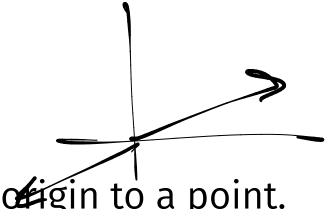
Lecture 6 | Part 1

Vectors

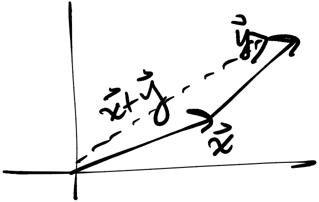
And now for something completely different...

- This and the next lecture will be linear algebra refreshers.
- Today: what is a matrix?
- Next lecture: what are eigenvectors/values?

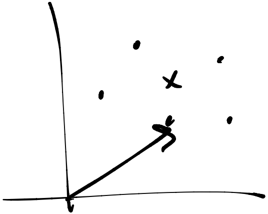
Vectors



- A vector $\vec{x}$ is an arrow from the origin to a point.
- We can make new arrows by:
  - scaling: $\alpha\vec{x}$
  - addition: $\vec{x} + \vec{y}$
  - both: $\alpha\vec{x} + \beta\vec{y}$



- $\|\vec{x}\|$ is the **norm** (or length) of $\vec{x}$



Linear Combinations

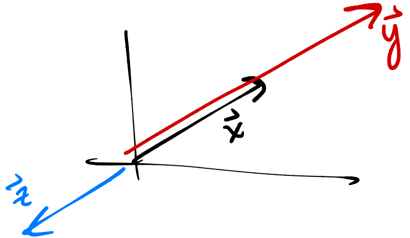
- We can add together a bunch of arrows:

$$\vec{y} = \alpha_1 \vec{x}^{(1)} + \alpha_2 \vec{x}^{(2)} + \ldots + \alpha_n \vec{x}^{(n)}$$

- This is a **linear combination** of $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$

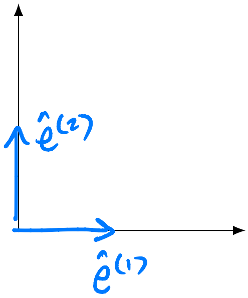
Parallel Vectors

- Two vectors $\vec{x}$ and $\vec{y}$ are **parallel** if (and only if) there is a scalar λ such that $\vec{x} = \lambda\vec{y}$.



Standard Basis Vectors

- $\hat{e}^{(1)}$ and $\hat{e}^{(2)}$ are the **standard basis vectors** in $\mathbb{R}^2$.
  - $\|\hat{e}^{(1)}\| = \|\hat{e}^{(2)}\| = 1$



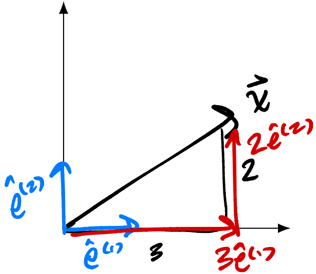
Standard Basis Vectors

- $\hat{e}^{(1)}, \dots, \hat{e}^{(d)}$ are the **standard basis vectors** in $\mathbb{R}^d$.

Decompositions

- We can **decompose** any vector $\vec{x} \in \mathbb{R}^2$ in terms of $\hat{e}^{(1)}$ and $\hat{e}^{(1)}$
  - Write: $\vec{x} = x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)}$

$$\vec{x} = 3\hat{e}^{(1)} + 2\hat{e}^{(2)}$$



Decompositions

- We can **decompose** any vector $\vec{x} \in \mathbb{R}^d$ in terms of $\hat{e}^{(1)}, \hat{e}^{(2)}, \dots, \hat{e}^{(d)}$
  - Write: $\vec{x} = x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)} + \dots + x_d \hat{e}^{(d)}$

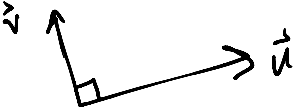
Coordinate Vectors

- We often write a vector $\vec{x}$ as a **coordinate vector**:

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

- Meaning: $\vec{x} = x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)} + \ldots + x_d \hat{e}^{(d)}$

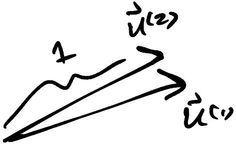
Dot Product



- The **dot product** of $\vec{u}$ and $\vec{v}$ is defined as:

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

where θ is the angle between $\vec{u}$ and $\vec{v}$.



- $\vec{u} \cdot \vec{v} = 0$ if and only if $\vec{u}$ and $\vec{v}$ are orthogonal

Dot Product (Coordinate Form)

- In terms of coordinate vectors:

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v}$$

$$= \begin{pmatrix} u_1 & u_2 & \cdots & u_d \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \cdots \\ v_d \end{pmatrix}$$

$$\|\vec{u}\| \cdot \|\vec{v}\| \cos\theta = u_1 v_1 + u_2 v_2 + \ldots + u_d v_d$$

1) $\vec{v} \cdot \vec{v} = \|\vec{v}\| \, \|\vec{v}\| \cos\theta = \|\vec{v}\|^2$

$\cos\theta = 1$ bec. $\theta = 0$

$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \dots + v_d^2}$

Exercise

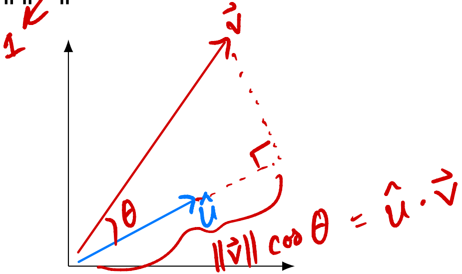
Show that $\vec{v} \cdot \vec{v} = \|\vec{v}\|^2$.

$$\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix}$$

$$\vec{v} \cdot \vec{v} = (v_1 \; v_2 \; \dots \; v_d) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} = v_1^2 + v_2^2 + \dots + v_d^2$$

Projections

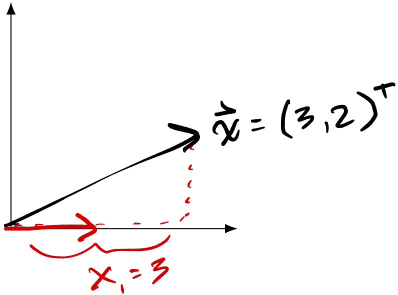
- If $\hat{u}$ is a unit vector, $\vec{v} \cdot \hat{u}$ is the "part of $\vec{v}$ that lies in the direction of $\hat{u}$".
  - $\vec{v} \cdot \hat{u} = \|\vec{v}\| \, \|\hat{u}\| \cos\theta$



Projections

$\vec{x} \cdot \hat{e}^{(1)}$

- Namely, if $\vec{x} = (x_1, \ldots, x_d)^T$, then $\vec{x} \cdot \hat{e}^{(k)} = x_k$.



DSC 190

Machine Learning: Representations

Lecture 6 | Part 2

Functions of a Vector

Functions of a Vector

- In ML, we often work with functions of a vector: $f : \mathbb{R}^d \to \mathbb{R}^{d'}$.
- Example: a prediction function, $H(\vec{x})$.
- Functions of a vector can return:
  - a number: $f : \mathbb{R}^d \to \mathbb{R}^1$
  - a vector $\vec{f} : \mathbb{R}^d \to \mathbb{R}^{d'}$
  - something else?

Transformations

- A **transformation** $\vec{f}$ is a function that takes in a vector, and returns a vector *of the same dimensionality*.
- That is, $\vec{f} : \mathbb{R}^d \to \mathbb{R}^d$.

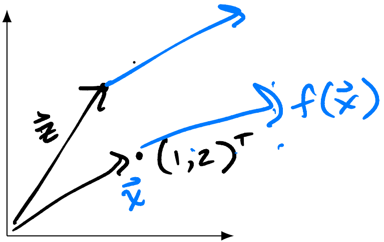
Visualizing Transformations

- A transformation is a **vector field**.
  - Assigns a vector to each point in space.
  - Example: $\vec{f}(\vec{x}) = (3x_1, x_2)^T$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

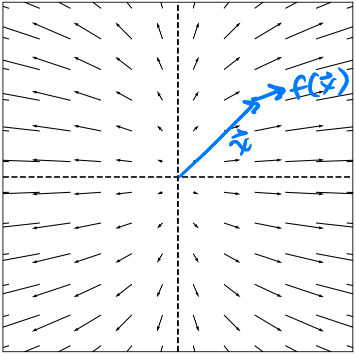
$$f((1,2)) = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

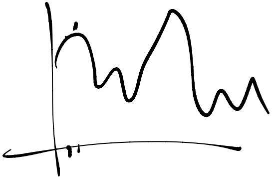
$$f((-5,0)) = \begin{pmatrix} -15 \\ 0 \end{pmatrix}$$

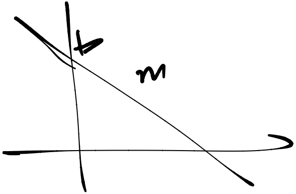


Example

- $\vec{f}(\vec{x}) = (3x_1, x_2)^T$

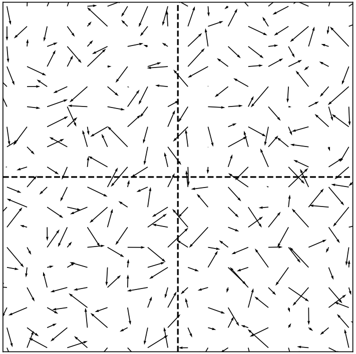






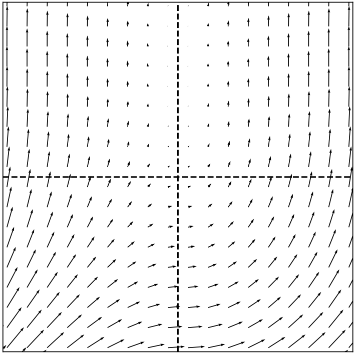
Arbitrary Transformations

- Arbitrary transformations can be quite complex.



Arbitrary Transformations

- Arbitrary transformations can be quite complex.



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

$$f(ax_1 + bx_2) = m(ax_1 + bx_2)$$

$$= amx_1 + bmx_2$$

$$= af(x_1) + bf(x_2)$$

Linear Transformations

- Luckily, we often[1] work with simpler, **linear transformations**.

- A transformation f is linear if:

$$\vec{f}(\alpha\vec{x} + \beta\vec{y}) = \alpha\vec{f}(\vec{x}) + \beta\vec{f}(\vec{y})$$

---

[1]Sometimes, just to make the math tractable!

Implications of Linearity

$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$\vec{x} = x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)}$

- Suppose $\vec{f}$ is a linear transformation. Then:

$$\begin{aligned} \vec{f}(\vec{x}) &= \vec{f}(x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)}) \\ &= x_1 \vec{f}(\hat{e}^{(1)}) + x_2 \vec{f}(\hat{e}^{(2)}) \end{aligned}$$

- I.e., $\vec{f}$ is **totally determined** by what it does to the basis vectors.

The **Complexity** of Arbitrary Transformations

- Suppose f is an **arbitrary** transformation.
- I tell you $\vec{f}(\hat{e}^{(1)}) = (2, 1)^T$ and $\vec{f}(\hat{e}^{(2)}) = (-3, 0)^T$.
- I tell you $\vec{x} = (x_1, x_2)^T$.
- What is $\vec{f}(\vec{x})$?

The **Simplicity** of Linear Transformations

- Suppose f is a **linear** transformation.
- I tell you $\vec{f}(\hat{e}^{(1)}) = (2, 1)^T$ and $\vec{f}(\hat{e}^{(2)}) = (-3, 0)^T$.
- I tell you $\vec{x} = (x_1, x_2)^T$.
- What is $\vec{f}(\vec{x})$?

$$\vec{f}(\vec{x}) = \vec{f}((3,-4)^T) = \vec{f}(3\hat{e}^{(1)} - 4\hat{e}^{(2)}) = 3f(\hat{e}^{(1)}) - 4f(\hat{e}^{(2)})$$

$$= 3\begin{pmatrix}2\\1\end{pmatrix} - 4\begin{pmatrix}-3\\0\end{pmatrix}$$

Exercise

- Suppose f is a **linear** transformation.
- I tell you $\vec{f}(\hat{e}^{(1)}) = (2, 1)^T$ and $\vec{f}(\hat{e}^{(2)}) = (-3, 0)^T$.
- I tell you $\vec{x} = (3, -4)^T$.
- What is $\vec{f}(\vec{x})$?

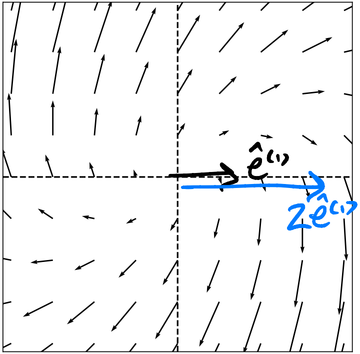
$$= \begin{pmatrix}6\\3\end{pmatrix} + \begin{pmatrix}12\\0\end{pmatrix} = \begin{pmatrix}18\\3\end{pmatrix}$$

Key Fact

- Linear functions are determined **entirely** by what they do on the basis vectors.
- I.e., to tell you what f does, I only need to tell you $\vec{f}(\hat{e}^{(1)})$ and $\vec{f}(\hat{e}^{(2)})$.
- This makes the math easy!

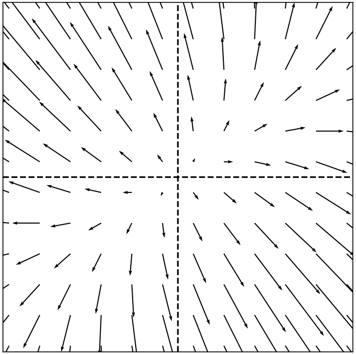
Example Linear Transformation

- $\vec{f}(\vec{x}) = (x_1 + 3x_2, -3x_1 + 5x_2)^T$ $\quad f(z\hat{e}^{(1)})$



Another Example Linear Transformation

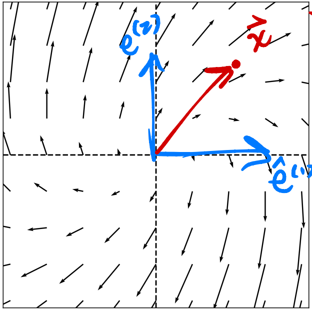
- $\vec{f}(\vec{x}) = (2x_1 - x_2, -x_1 + 3x_2)^T$

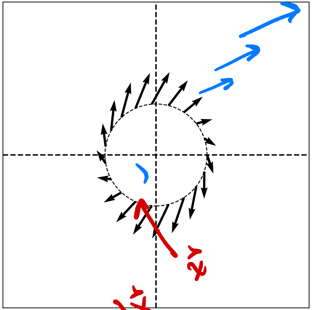


Note

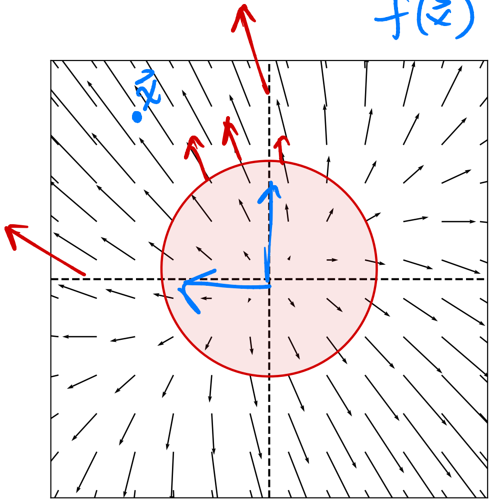
- Because of linearity, along any given direction $\vec{f}$ changes only in scale.

$$\vec{f}(\lambda \hat{x}) = \lambda \vec{f}(\hat{x})$$

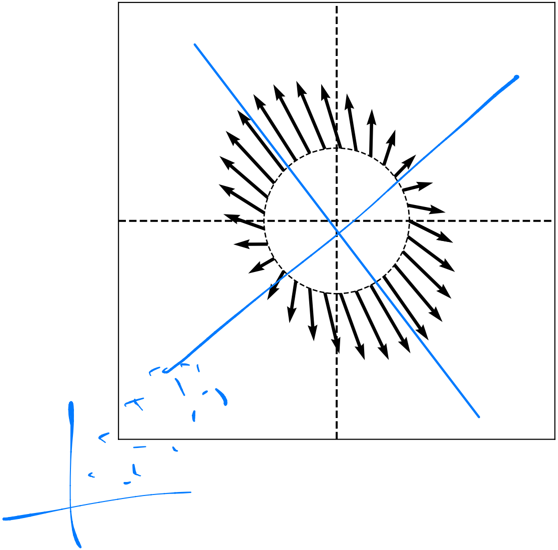




$f(\vec{x})$



$f(a\vec{x}+b\vec{y}) = af(\vec{x})+bf(\vec{y})$



DSC 190

Machine Learning: Representations

Lecture 6 | Part 3

Matrices

Matrices?

- I thought this was supposed to be about linear algebra... Where are the matrices?

Matrices?

- I thought this was supposed to be about linear algebra... Where are the matrices?
- What is a matrix, anyways?

What is a matrix?

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

What is matrix multiplication?

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

$$= \begin{pmatrix} -2 + 2 + 9 \\ \vdots \end{pmatrix} = \begin{pmatrix} 9 \\ \vdots \end{pmatrix}$$

A low-level definition

$$(A\vec{x})_i = \sum_{j=1}^{n} A_{ij} x_j$$

A low-level interpretation

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

In general...

$$\begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{a}^{(1)} & \vec{a}^{(2)} & \vec{a}^{(3)} \\ \downarrow & \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1\vec{a}^{(1)} + x_2\vec{a}^{(2)} + x_3\vec{a}^{(3)}$$

What are they, *really*?

- Matrices are sometimes just tables of numbers.
- But they often have a deeper meaning.

Main Idea

A square $(n \times n)$ matrix can be interpreted as a compact representation of a linear transformation $\vec{f} : \mathbb{R}^n \to \mathbb{R}^n$.

What's more, if A represents $\vec{f}$, then $A\vec{x} = \vec{f}(\vec{x})$; that is, multiplying by A is the same as evaluating $\vec{f}$.

Recall: Linear Transformations

- A **transformation** $\vec{f}(\vec{x})$ is a function which takes a vector as input and returns a vector of the same dimensionality.

- A transformation f is **linear** if

$$\vec{f}(\alpha\vec{u} + \beta\vec{v}) = \alpha\vec{f}(\vec{u}) + \beta\vec{f}(\vec{v})$$

Recall: Linear Transformations

- A **key** property: to compute $\vec{f}(\vec{x})$, we only need to know what f does to basis vectors.

- Example:

$$\vec{x} = 3\hat{e}^{(1)} - 4\hat{e}^{(2)} = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$$

$$\vec{f}(\hat{e}^{(1)}) = -\hat{e}^{(1)} + 3\hat{e}^{(2)}$$

$$\vec{f}(\hat{e}^{(2)}) = 2\hat{e}^{(1)}$$

$$\vec{f}(\vec{x}) =$$

Matrices

- $\vec{f}$ defined by what it does to basis vectors
- Place $\vec{f}(\hat{e}^{(1)}), \vec{f}(\hat{e}^{(2)}), \ldots$ into a table as columns
- This is the **matrix** representing[2] f

$$\vec{f}(\hat{e}^{(1)}) = -\hat{e}^{(1)} + 3\hat{e}^{(2)} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

$$\vec{f}(\hat{e}^{(2)}) = 2\hat{e}^{(1)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$A = \begin{pmatrix} -1 & 2 \\ 3 & 0 \end{pmatrix}$$

[2]with respect to the basis $\hat{e}^{(1)}, \hat{e}^{(2)}$

Example

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

$$\vec{f}(\hat{e}^{(1)}) = (1, 4, 7)^T$$

$$\vec{f}(\hat{e}^{(2)}) = (2, 5, 7)^T$$

$$\vec{f}(\hat{e}^{(3)}) = (3, 6, 9)^T$$

Main Idea

A square $(n \times n)$ matrix can be interpreted as a compact representation of a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$.

Matrix Multiplication

- Matrix A represents a function f
- Matrix multiplication $A\vec{x}$ **evaluates** $\vec{f}(\vec{x})$

Matrix Multiplication

$$\vec{x} = x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)} + x_3 \hat{e}^{(3)} = (x_1, x_2, x_3)^T$$

$$\vec{f}(\vec{x}) = x_1 \vec{f}(\hat{e}^{(1)}) + x_2 \vec{f}(\hat{e}^{(2)}) + x_3 \vec{f}(\hat{e}^{(3)})$$

$$A = \begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{f}(\hat{e}^{(1)}) & \vec{f}(\hat{e}^{(2)}) & \vec{f}(\hat{e}^{(3)}) \\ \downarrow & \downarrow & \downarrow \end{pmatrix}$$

$$A\vec{x} = \begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{f}(\hat{e}^{(1)}) & \vec{f}(\hat{e}^{(2)}) & \vec{f}(\hat{e}^{(3)}) \\ \downarrow & \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$= x_1 \vec{f}(\hat{e}^{(1)}) + x_2 \vec{f}(\hat{e}^{(2)}) + x_3 \vec{f}(\hat{e}^{(3)})$$

Example

$$\vec{x} = 3\hat{e}^{(1)} - 4\hat{e}^{(2)} = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$$

$$\vec{f}(\hat{e}^{(1)}) = -\hat{e}^{(1)} + 3\hat{e}^{(2)}$$

$$\vec{f}(\hat{e}^{(2)}) = 2\hat{e}^{(1)}$$

$$\vec{f}(\vec{x}) = f\left(3e^{(1)} - 4e^{(2)}\right)$$

$$= 3f\left(e^{(1)}\right) - 4f\left(e^{(2)}\right)$$

$$= 3\begin{pmatrix} -1 \\ 3 \end{pmatrix} - 4\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ 9 \end{pmatrix} + \begin{pmatrix} -8 \\ 0 \end{pmatrix} \longrightarrow = \begin{pmatrix} -11 \\ 9 \end{pmatrix}$$

$$A = \begin{pmatrix} -1 & 2 \\ 3 & 0 \end{pmatrix}$$

$$A\vec{x} = \begin{pmatrix} -1 & 2 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} 3 \\ -4 \end{pmatrix} = \begin{pmatrix} -3 - 8 \\ 9 + 0 \end{pmatrix}$$

Main Idea

A square $(n \times n)$ matrix can be interpreted as a compact representation of a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$. Matrix multiplication with a vector $\vec{x}$ evaluates $\vec{f}(\vec{x})$.

Note

- All of this works because we assumed $\vec{f}$ is **linear**.
- If it isn't, evaluating $\vec{f}$ isn't so simple.

Note

- All of this works because we assumed $\vec{f}$ is **linear**.
- If it isn't, evaluating $\vec{f}$ isn't so simple.
- Linear algebra = simple!