we'll start @ 11:47

As you get settled...

- Get out your notes
- Get out a place to do today's ICA (5)
- Where are you on HW 1?

  A. I haven't looked at it

  B. I've glanced at the problems

  C. I've gotten started but I'm not very far

  D. I'm probably half way through

  E. I'm finished/almost finished

Now playing:
"When We were young",
Adele

YOASOBI, this is
a single from 2019 w/
a title in japanese

Linear Perceptrons

Given the features of **snout length** and **fluffiness**, featurize the following data points:

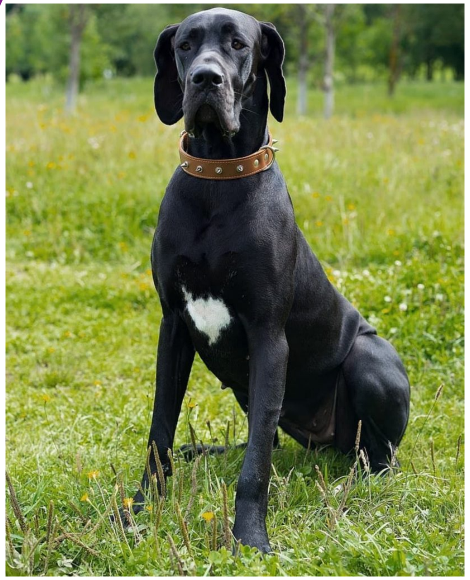
a)

$$\begin{bmatrix} 3 \\ 10 \end{bmatrix}$$

b)

$$\begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

c)



$$\begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

d)

$$\begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Machine Learning

- All machine learning models that do classification have the following components:
  ↳ is it a dog?

- A way to **represent the input data** -> last lecture → featurize into num. vectors

- A **classification function** -> our starting point today

- A way to **train the model** -> also today! → "learning"

Linear Perceptron

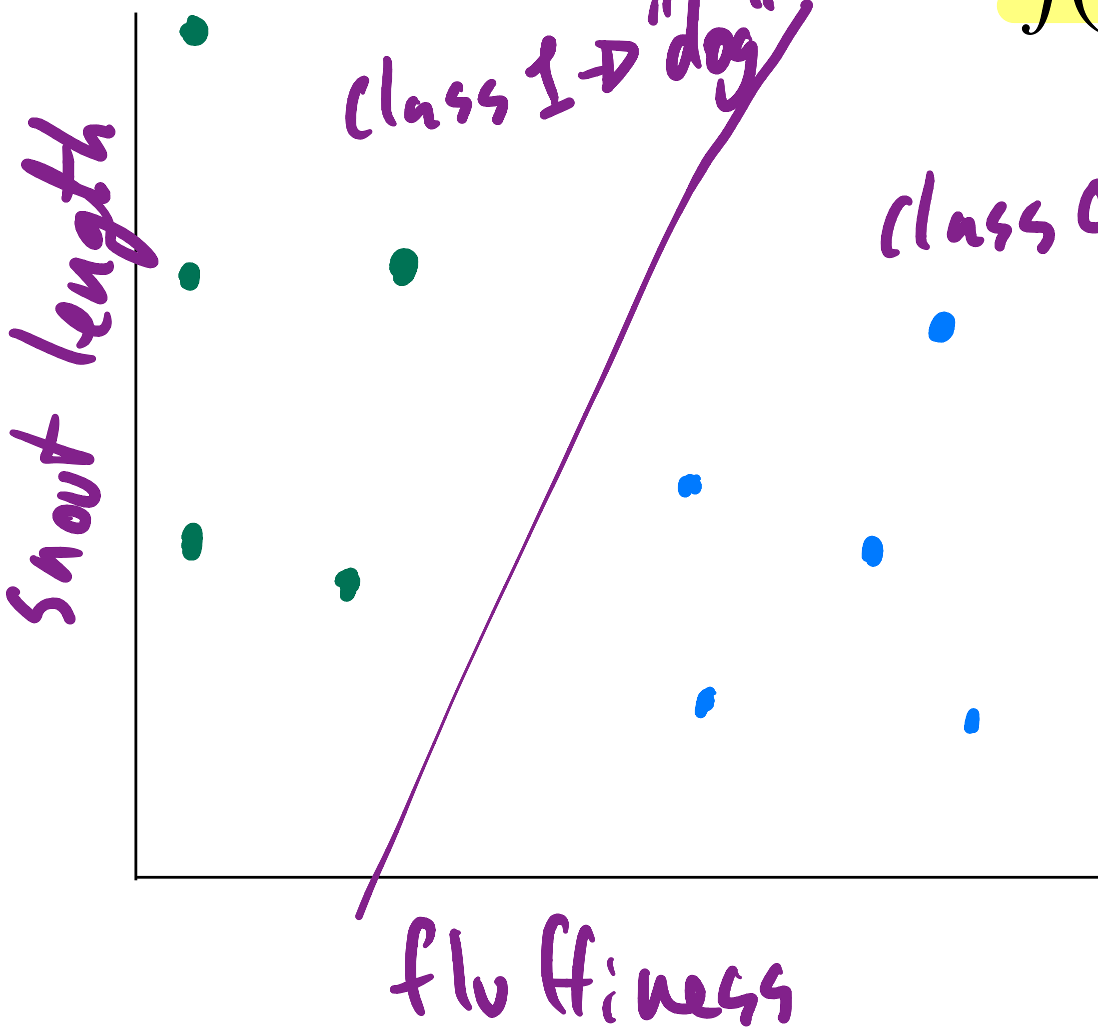
- A function that estimates one of two classes (a **binary classifier**), defined by the vector $\vec{w}$ → "weights"

$$f(\vec{x}) = 1 \textbf{ if } \vec{x} \cdot \vec{w} \geq 0 \textbf{ else } 0$$

class 1 → "dog"

class 0 → "cat"

→ new data point, take dot prod, assign a class



Linear Perceptron

ICA Question 1: what is a linear perceptron's class estimate for the following samples if the perceptron is defined by the vector:

$$f(\vec{x}) = 1 \textbf{ if } \vec{x} \cdot \vec{w} \geq 0 \textbf{ else } 0$$

dog

squash face cat

~~fluffy puppy~~

snout length

fluffiness

$$\vec{w} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \qquad \vec{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \vec{x}_2 = \begin{bmatrix} 0 \\ 4 \end{bmatrix} \qquad \vec{x}_3 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$\vec{w} \cdot \vec{x}_1 = 1$ $\quad \vec{w} \cdot \vec{x}_2 = -8$ $\quad \vec{w} \cdot \vec{x}_3 = -7$

class: 1 $\quad$ class: 0 $\quad$ class: $\boxed{0}$

not correct

Linear Perceptron

ICA Question 1: what is a linear perceptron's class estimate for the following samples if the perceptron is defined by the vector:

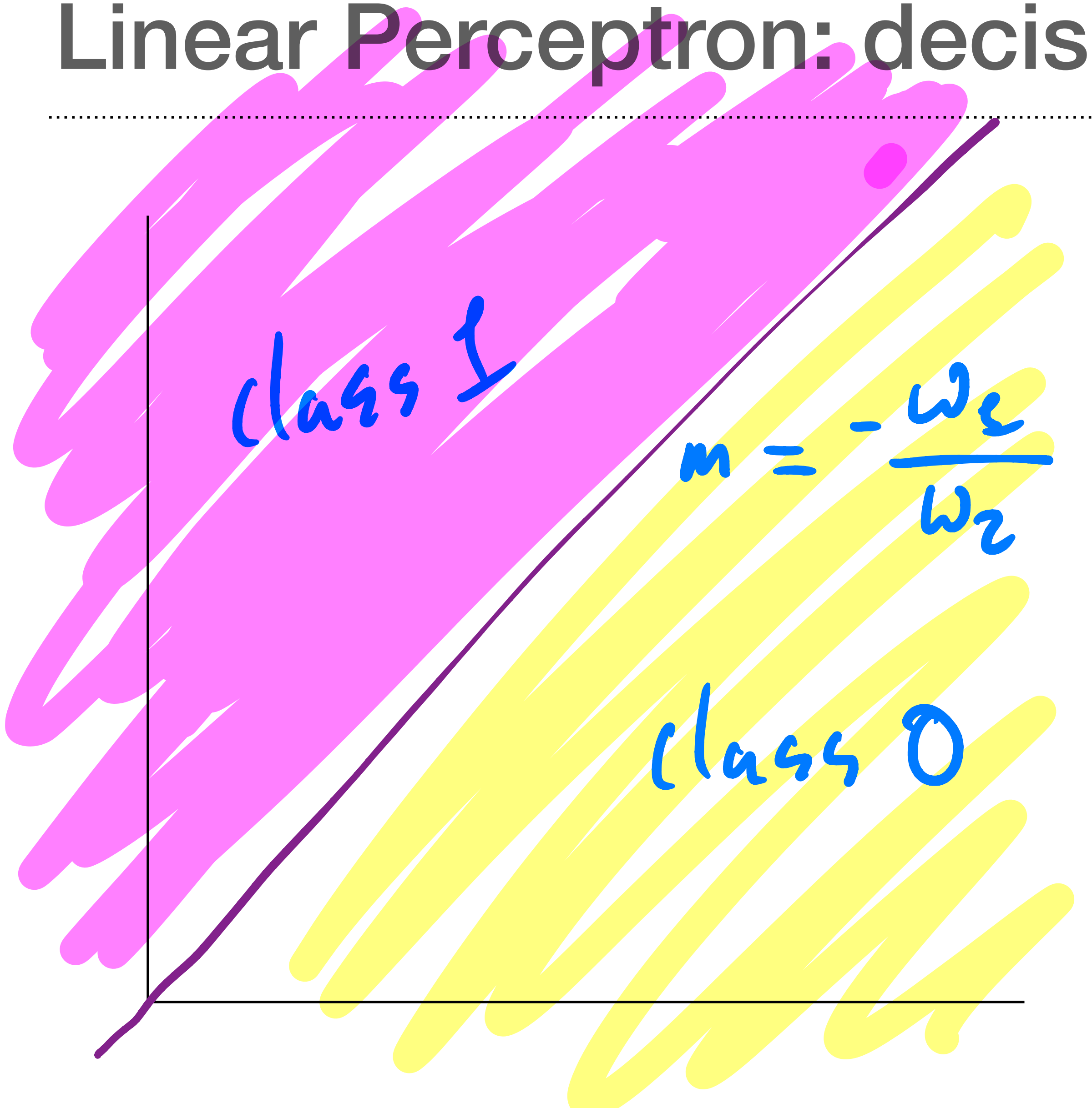
$$f(\overrightarrow{x}) = 1 \textbf{ if } \overrightarrow{x} \cdot \overrightarrow{w} \geq 0 \textbf{ else } 0$$

$$\overrightarrow{w} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

 $= \begin{bmatrix} 1 \\ 4 \end{bmatrix}$

! classification error → will want to update our weights!

Linear Perceptron: decision boundary



- Visualize all points where are estimated to be either class 1 or class 0

- $\vec{x} \cdot \vec{w} = 0$

$$= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = 0$$

$$= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 0$$

$$= x_1 w_1 + x_2 w_2 = 0$$

$$x_2 = -\frac{w_1}{w_2} x_1$$

equivalent, 0-indexed way to write this

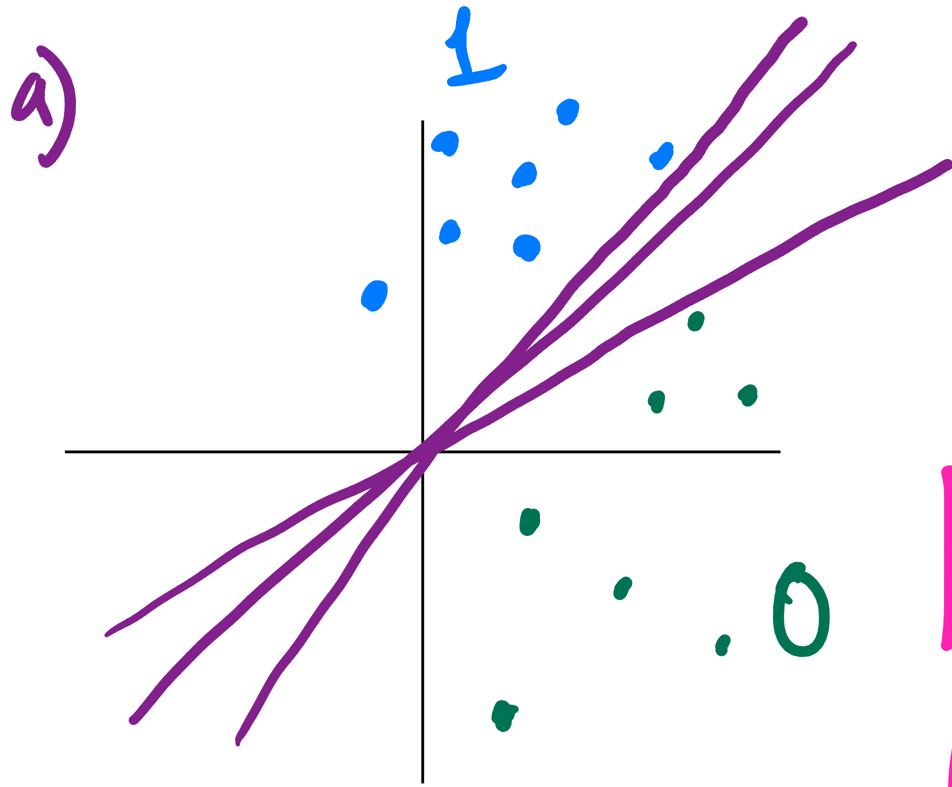
$$\boxed{x_1 = -\frac{w_0}{w_1} x_0}$$

Linear Perceptron: decision boundary

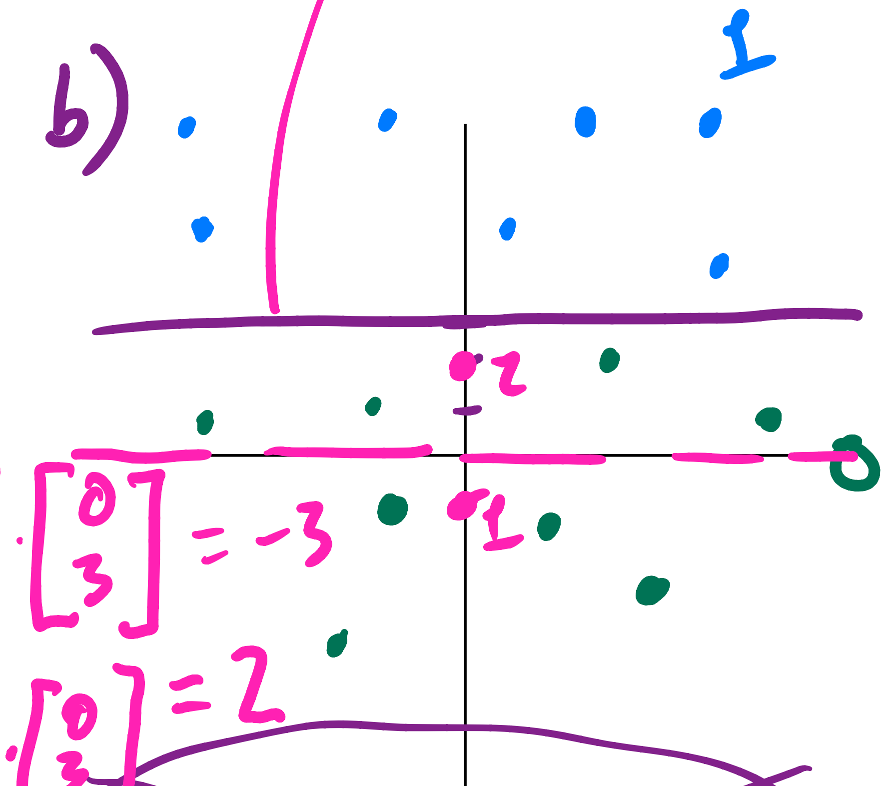
but the slope is 0!

ICA Question 2: find a perceptron weight vector $\vec{w}$ that distinguishes all samples below. Double check your weights with a sample from each class!

a)



b)



$\begin{bmatrix} 0 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 3 \end{bmatrix} = -3$

$\begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 3 \end{bmatrix} = 2$

A: w for a

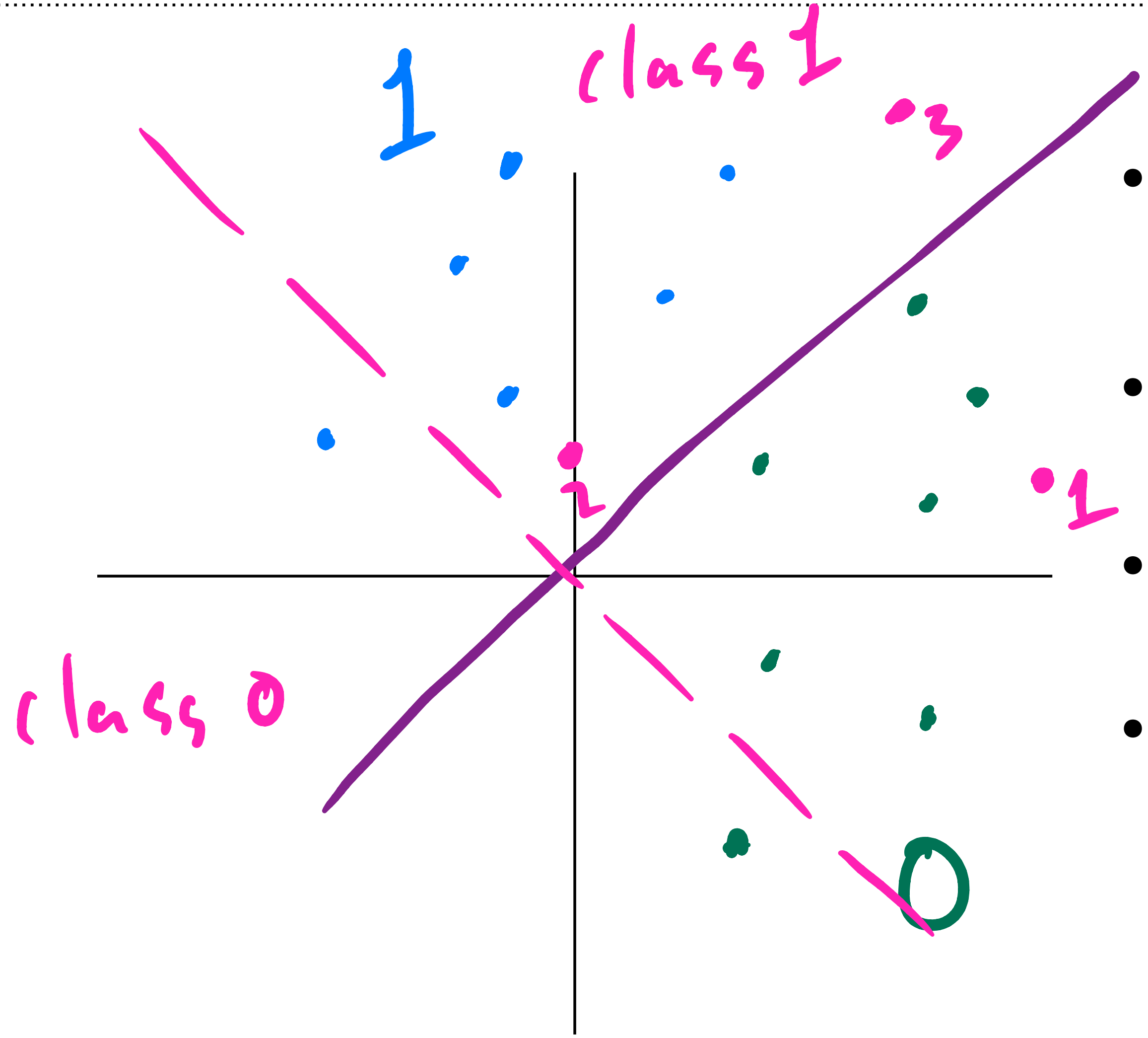
B: w for b

C: none of them

$\vec{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$\vec{w} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$ ~~$m = -\frac{0}{3}$~~

Linear Perceptron: ICA 2



- $m = \dfrac{-w_0}{w_1}$
- $\vec{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- Check: $\vec{x}_1 \cdot \vec{w} = \begin{bmatrix} 10 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 11$
- Estimated class is: $\underline{\underline{1}}$ → we wanted 0

$$\vec{x}_2 \cdot \vec{w} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 + 2 = 2$$

$$x_3 \cdot \vec{w} = \begin{bmatrix} 8 \\ 10 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 8 + 10 = 18$$

Linear Perceptron: ICA 2

class 1 over here

w/ updated weights class 0

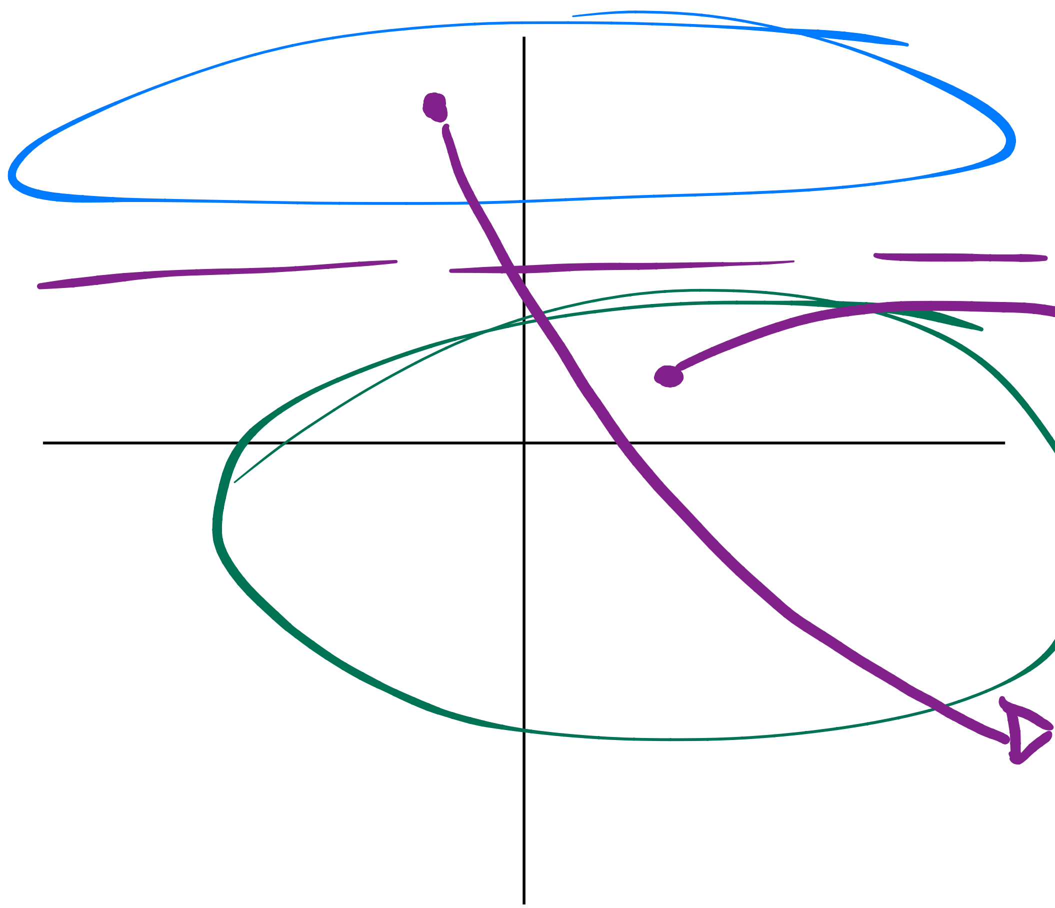
- $m = \dfrac{-w_0}{w_1}$
- $\vec{w} = \begin{bmatrix} -1 \\ +1 \end{bmatrix}$
- Check: $\vec{x} \cdot \vec{w} = \begin{bmatrix} 10 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ +1 \end{bmatrix} = \underline{-9}$ 0
- Estimated class is: $\underline{1}$ wanted 0

flip all signs, switches the decision boundary

$\vec{x}_2 \cdot \vec{w} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ +1 \end{bmatrix} = +2$

↳ class is 1 wanted 0

Linear Perceptron: bias term



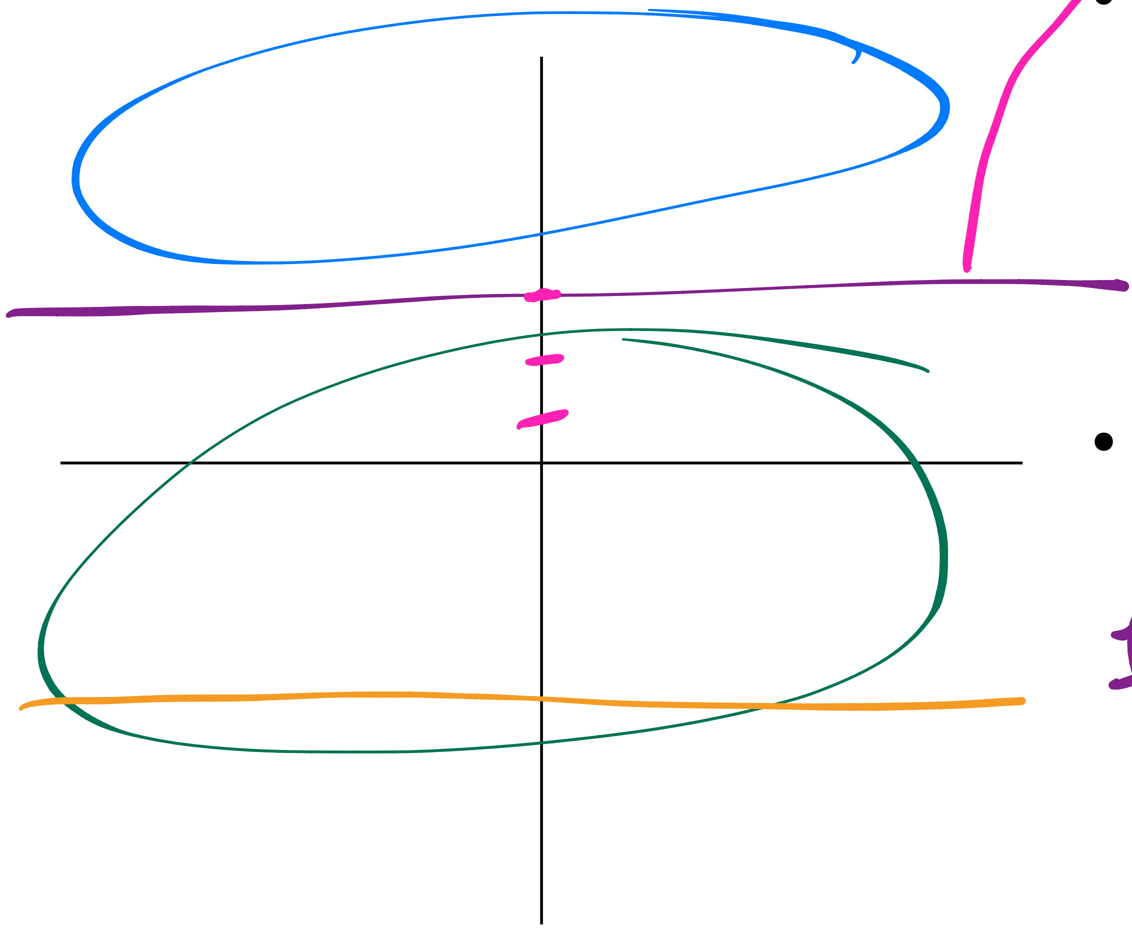
- The **bias** term adds a "fake feature" whose value is **always 1**
- $\vec{x} = \begin{bmatrix}1\\4\\3\end{bmatrix}$
- $\vec{y} = \begin{bmatrix}1\\-2\\6\end{bmatrix}$

- no line through the origin separates these points!

$y = mx + b$

Linear Perceptron: bias term

$$\begin{bmatrix} -9 \\ 0 \\ 3 \end{bmatrix} \approx \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$



- recall: boundary is all points w/ $\vec{x} \cdot \vec{w} = 0$

bias term, always 1

- $$\begin{bmatrix} 1 \\ \cancel{x_1} \\ \cancel{x_2} \end{bmatrix} \cdot \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = 0$$

$$1 * w_0 + x_1 w_1 + x_2 w_2 = 0$$

$$x_2 = -\frac{w_0}{w_2} - \frac{w_1}{w_2} x_1$$

b/intercept m/slope

$$x_2 = b + m x_1$$
$$= m x_1 + b$$

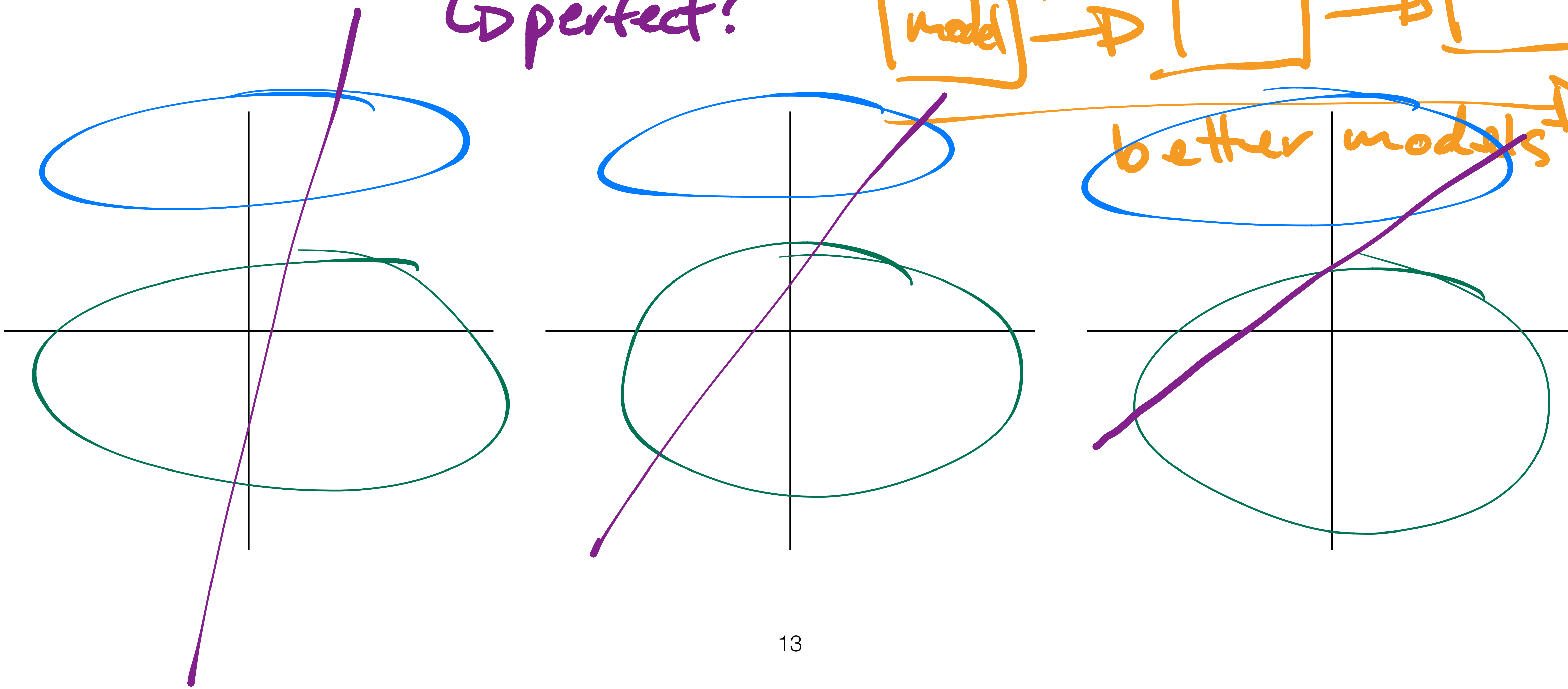
Linear Perceptron: Training

break: 12:50
↳ you don't have more ICA Qs today, so if you want to turn in now you can

- how do we train the model?
- how do we know when to stop training?

↳ perfect?

model $\xrightarrow{\text{train}}$ ☐ → ☐ → better models



Linear Perceptron: Training

- Aside: length & dot product
- $||\vec{x}|| = \sqrt{\sum_i x_i^2}$
- $\vec{x} \cdot \vec{x} = \begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix} = x_0^2 + x_1^2 + \ldots\ldots + x_n^2$

  $-3 \cdot -3 = 9$
- $\vec{x} \cdot \vec{x} = ||\vec{x}||^2$
- Note: $\vec{x} \cdot \vec{x}$ is ____________ positive

  Always

  Sometimes

  never

Linear Perceptron: Training

- Aside: dot product is **linear**

$$f(x+y) \qquad\qquad f(x)+f(y)$$

- $\vec{x}\cdot(\vec{y}+\vec{z}) = \vec{x}\cdot\vec{y}+\vec{x}\cdot\vec{z}$

$$\begin{bmatrix}1\\2\end{bmatrix}\cdot\left(\begin{bmatrix}3\\4\end{bmatrix}+\begin{bmatrix}5\\6\end{bmatrix}\right) = \begin{bmatrix}1\\2\end{bmatrix}\cdot\begin{bmatrix}3\\4\end{bmatrix}+\begin{bmatrix}1\\2\end{bmatrix}\cdot\begin{bmatrix}5\\6\end{bmatrix}$$

$$\begin{bmatrix}1\\2\end{bmatrix}\cdot\left(\begin{bmatrix}8\\10\end{bmatrix}\right) = 3+8+5+12$$

$$8+20 = 11+17$$

$$28 = 28$$

Linear Perceptron: Training

- Suppose $\overrightarrow{x}$ belongs to class 1 but our current perceptron estimates it as class 0 → our guess is negative
  - we want $\overrightarrow{x} \cdot \overrightarrow{w}$ to be **larger**
- We'll **add** the value of $\overrightarrow{x}$ to $\overrightarrow{w}$:
  - $\overrightarrow{w}' = \overrightarrow{w} + \overrightarrow{x}$
    ↳ updated weights
  - $\overrightarrow{x} \cdot \overrightarrow{w}' = \overrightarrow{x} \cdot (\overrightarrow{w} + \overrightarrow{x}) = \vec{x} \cdot \vec{w} + \vec{x} \cdot \vec{x}$
    ↳ larger than $\vec{x} \cdot \vec{w}$?
    ↳ yes!
    ↳ old dot prod.
    ↳ positive

Linear Perceptron: Training

- Suppose $\overrightarrow{x}$ belongs to class 0 but our current perceptron estimates it as class 1

    - we want $\overrightarrow{x} \cdot \overrightarrow{w}$ to be **smaller**

- We'll subtract the value of $\overrightarrow{x}$ ~~to~~ from $\overrightarrow{w}$:

    - $\overrightarrow{w}' = \vec{w} - \vec{x}$

Linear Perceptron: Training

- choose starting $\vec{w}$ arbitrarily
- for each $\vec{x}, y$: → one example datapoint, $\vec{x}$ is features, y is the label/class
  - $\hat{y} = \vec{w} \cdot \vec{x} \geq 0$ → current estimated class
  - if $\hat{y} = 0$ and $y = 1$
    - $\vec{w}\,' = \vec{w}$ $+ \alpha \vec{x}$
      - α → learning rate
  - else if $\hat{y} = 1$ and $y = 0$
    - $\vec{w}\,' = \vec{w}$ $- \alpha \vec{x}$
  - else (we're correct)
    - ~~???~~ → nothing

Linear Perceptron: Learning rate

- The **learning rate** helps our model become more robust to wild swings.
- Imagine that you had two different training samples:

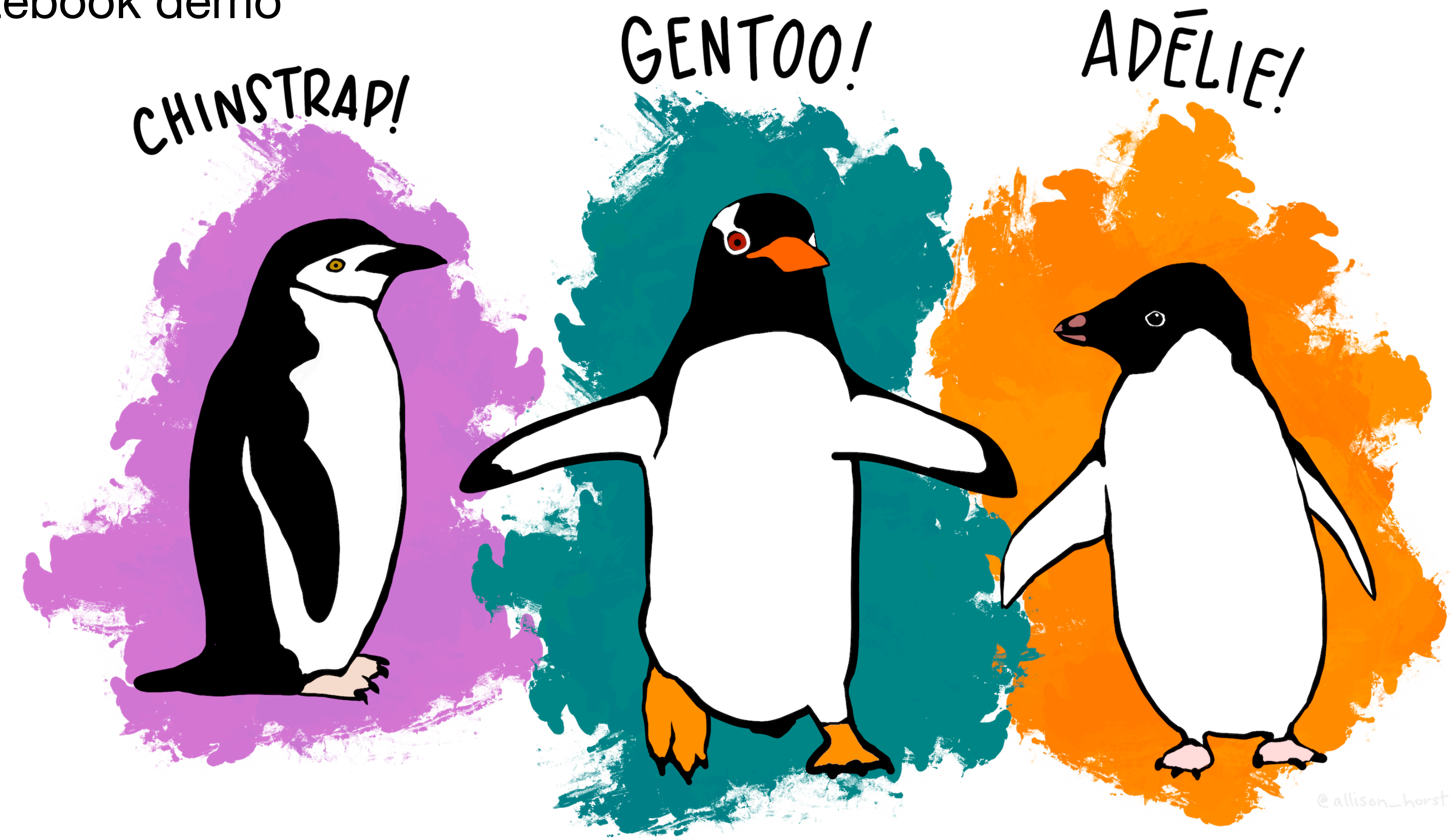
$$\vec{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \vec{x}_2 = \begin{bmatrix} 1000 \\ 3000 \end{bmatrix}$$

↳ multiply by a # in range $(0, 1]$

↳ experimentally determined

Linear Perceptron: demo

- Notebook demo



Schedule

wow, so many office hours now!

khouryofficehours.com

Turn in ICA 5 on Gradescope

HW 1 is due on Wednesday!

We are remote until Feb 5th. Next Monday we'll be in person in Snell Engineering 108.

Mon	Tue	Wed	Thu	Fri	Sat	Sun
January 31st Lecture 5 - Linear Perceptron	**Felix OH Calendly**	**HW 1 due @ 11:59pm**	Lecture 6 - matrix multiplication, transforms **Felix OH Khoury Office Hours** **HW 2 released**			
February 7th Lecture 7 - Vector spaces **in Snell Engineering 108**	**Felix OH Calendly**		Lecture 8 - line of best fit **Felix OH Khoury Office Hours**			**HW 2 due @ 11:59pm**

HW 1: proving linearity

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

not linear: one counter example is sufficient

$$f(x) = |x|$$

P3: $\begin{bmatrix} \text{variables} \end{bmatrix} = \begin{bmatrix} \text{start} \end{bmatrix} + \begin{bmatrix} \text{coefficients} \\ \text{times} \\ \text{free variables} \end{bmatrix}$