As you walk in

Welcome!

1) Sit at a table next to another student
2) Make name plate
   - Fold paper in half
   - Write preferred name

Mathematical Foundations for Machine Learning

Linear Systems

Instructor: Pat Virtue

Today

Linear Systems

- Systems of equations
- Fitting linear models to data

Detailed course topics doc

https://docs.google.com/spreadsheets/d/1I8g9pR-krQNKVuiPrxGUEkSwdyDjBIpTr9inq6Dl8-U

Exercise

Given the following system of equations, *how* would you solve for acceptable values for $\theta_1, \theta_2, \theta_3$? (Solve it too ☺)

Alien coins! Your friend E.T. is helping you to learn alien currency. There are three different types of coins that have values $\theta_1, \theta_2, \theta_3$. There are three different piles of coins. E.T. is kind enough to tell us the total value of each of the three piles.

$$
\begin{aligned}
3\theta_1 + 0\theta_2 + 5\theta_3 &= 36 \\
4\theta_1 + 3\theta_2 + 2\theta_3 &= 46 \\
2\theta_1 + 2\theta_2 + 1\theta_3 &= 25
\end{aligned}
$$

Exercise

Given the following system of equations, how would you solve for acceptable values for $\theta_1, \theta_2, \theta_3$?

$$3\theta_1 + 0\theta_2 + 5\theta_3 = 36$$
$$4\theta_1 + 3\theta_2 + 2\theta_3 = 46$$
$$2\theta_1 + 2\theta_2 + 1\theta_3 = 25$$

Reduced echelon (row) form

Gaussian elimination

$\theta_1 = 7$

$\theta_2 = 4$

$\theta_2 = 3$

Exercise

Given the following system of equations, how would you solve for acceptable values for $\theta_1, \theta_2, \theta_3$?

$$3\theta_1 + 0\theta_2 + 5\theta_3 = 36$$
$$4\theta_1 + 3\theta_2 + 2\theta_3 = 46$$
$$2\theta_1 + 2\theta_2 + 1\theta_3 = 25$$

$$A = \begin{bmatrix} 3 & 0 & 5 \\ 4 & 3 & 2 \\ 2 & 2 & 1 \end{bmatrix} \quad \vec{b} = \begin{bmatrix} 36 \\ 46 \\ 25 \end{bmatrix}$$

$$\vec{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

$$A\theta = b$$
$$A^{-1}A\theta = A^{-1}b$$
$$\theta = A^{-1}b$$

Linear Algebra

Notation alert!

Linear algebra allows us to represent and operate upon sets of linear equations.

$$3\theta_1 + 0\theta_2 + 5\theta_3 = 36$$
$$4\theta_1 + 3\theta_2 + 2\theta_3 = 46$$
$$2\theta_1 + 2\theta_2 + 1\theta_3 = 25$$

$$V\boldsymbol{\theta} = \boldsymbol{u} \qquad V = \begin{bmatrix} 3 & 0 & 5 \\ 4 & 3 & 2 \\ 2 & 2 & 1 \end{bmatrix} \quad \boldsymbol{u} = \begin{bmatrix} 36 \\ 46 \\ 25 \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

Possible way to solve: $\boldsymbol{\theta} = V^{-1}\boldsymbol{u}$

(Actually a pretty unstable way to solve in general. More on this later.)

Poll 1

What could happen if we added one more equation?

$$3\theta_1 + 0\theta_2 + 5\theta_3 = 36$$
$$4\theta_1 + 3\theta_2 + 2\theta_3 = 46$$
$$2\theta_1 + 2\theta_2 + 1\theta_3 = 25$$

Select ALL that apply

A. No solution

B. One solution

C. Two solutions

D. Infinite solutions

E. Error

Poll 1

What could happen if we added one more equation?

$$3\theta_1 + 0\theta_2 + 5\theta_3 = 36$$
$$4\theta_1 + 3\theta_2 + 2\theta_3 = 46$$
$$2\theta_1 + 2\theta_2 + 1\theta_3 = 25$$

Overdetermined

Inconsistent

Select ALL that apply

✓ A. No solution 841

✓ B. One solution 72

✗ C. Two solutions

✗ D. Infinite solutions

~ E. Error

15-30

Sol for orig 3 works for 4th

or not

Poll 2

What could happen if we remove one of these equations?

$$3\theta_1 + 0\theta_2 + 5\theta_3 = 36$$
$$4\theta_1 + 3\theta_2 + 2\theta_3 = 46$$
$$2\theta_1 + 2\theta_2 + 1\theta_3 = 25$$

Select ALL that apply

A. No solution

B. One solution

C. Two solutions

D. Infinite solutions

E. Error

Poll 2

What could happen if we remove one of these equations?

$$3\theta_1 + 0\theta_2 + 5\theta_3 = 36$$
$$4\theta_1 + 3\theta_2 + 2\theta_3 = 46$$
$$\cancel{2\theta_1 + 2\theta_2 + 1\theta_3 = 25}$$

Select ALL that apply

A. No solution 16

B. One solution 4

C. Two solutions ×

D. Infinite solutions 96

E. Error ×

Underdetermined

Notation alert!

Exercise

Write a system of equations to help us fit a line to the following data:

$\mathcal{D} = \left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{N}$, where N is the number of points in the dataset

Specifically:

$$
\begin{aligned}
\mathcal{D} = \{&(-1,\ \ 2.5),\\
&(7,\ -1.5),\\
&(-5,\ 4.5),\\
&(1.5,\ 1.25)\}
\end{aligned}
$$

Notation alert!

Exercise

Write a system of equations to help us fit a line to the following data:

$\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$, where N is the number of points in the dataset

Specifically:

$$\mathcal{D} = \{(-1,\ 2.5),\ (7,\ -1.5),\ (-5,\ 4.5),\ (1.5,\ 1.25)\}$$

$(x^{(1)}, y^{(1)})$

$(x^{(2)}, y^{(2)})$

$\vdots$

$(x^{(N)}, y^{(N)})$

w → m → $ax + b = y$

$a(-1) + b = 2{,}5$

$a \cdot 7 + b = -1{,}5$

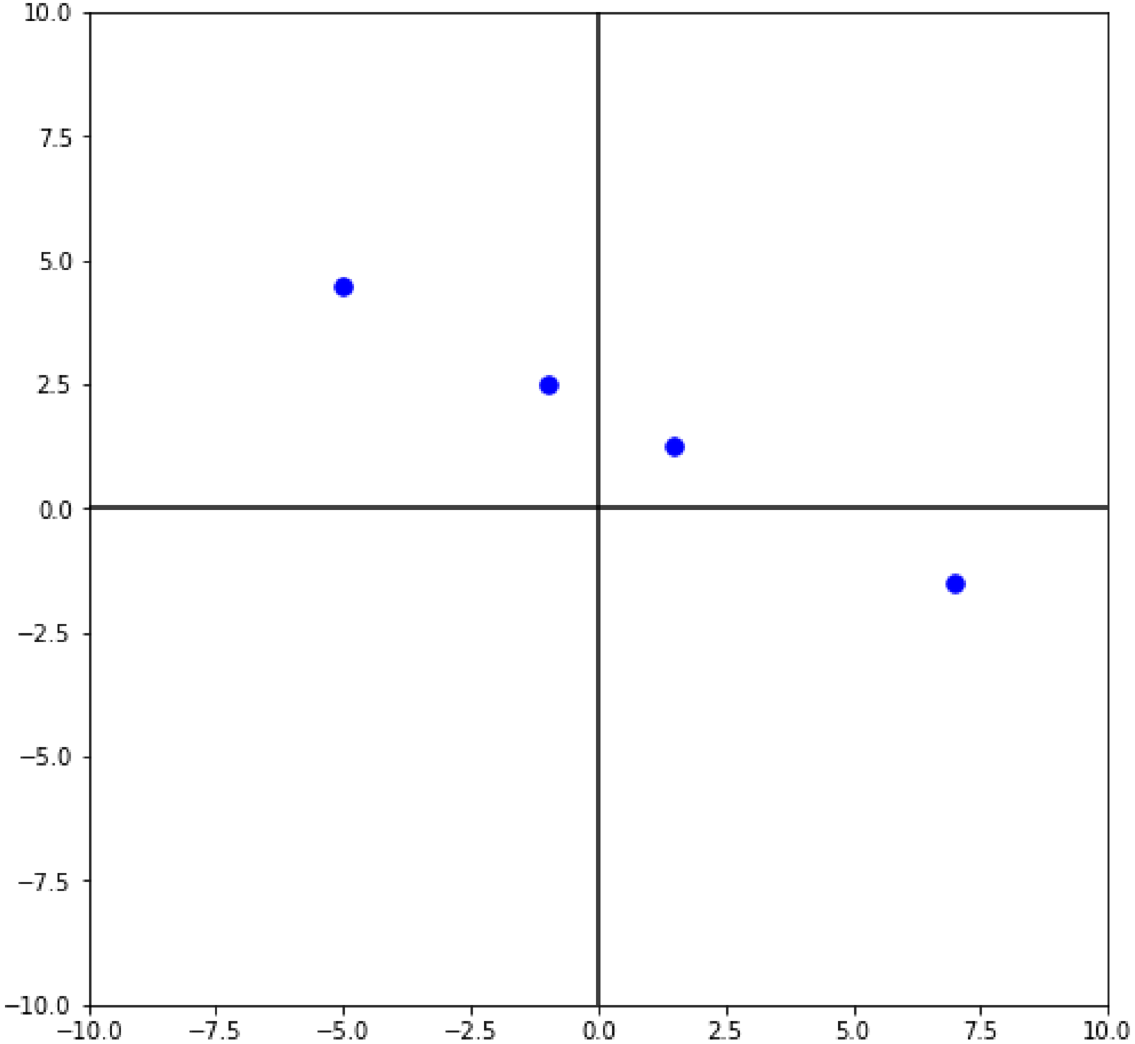
$a(-5) + b = 4{,}5$

$a(1{,}5) + b = 1{,}25$

Poll 3

Can we fit a linear model to this data?

$$\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{4}$$
$$= \{(-1,\ \ 2.5),$$
$$(7,\ -1.5),$$
$$(-5,\ 4.5),$$
$$(1.5,\ 1.25)\}$$



Poll 3

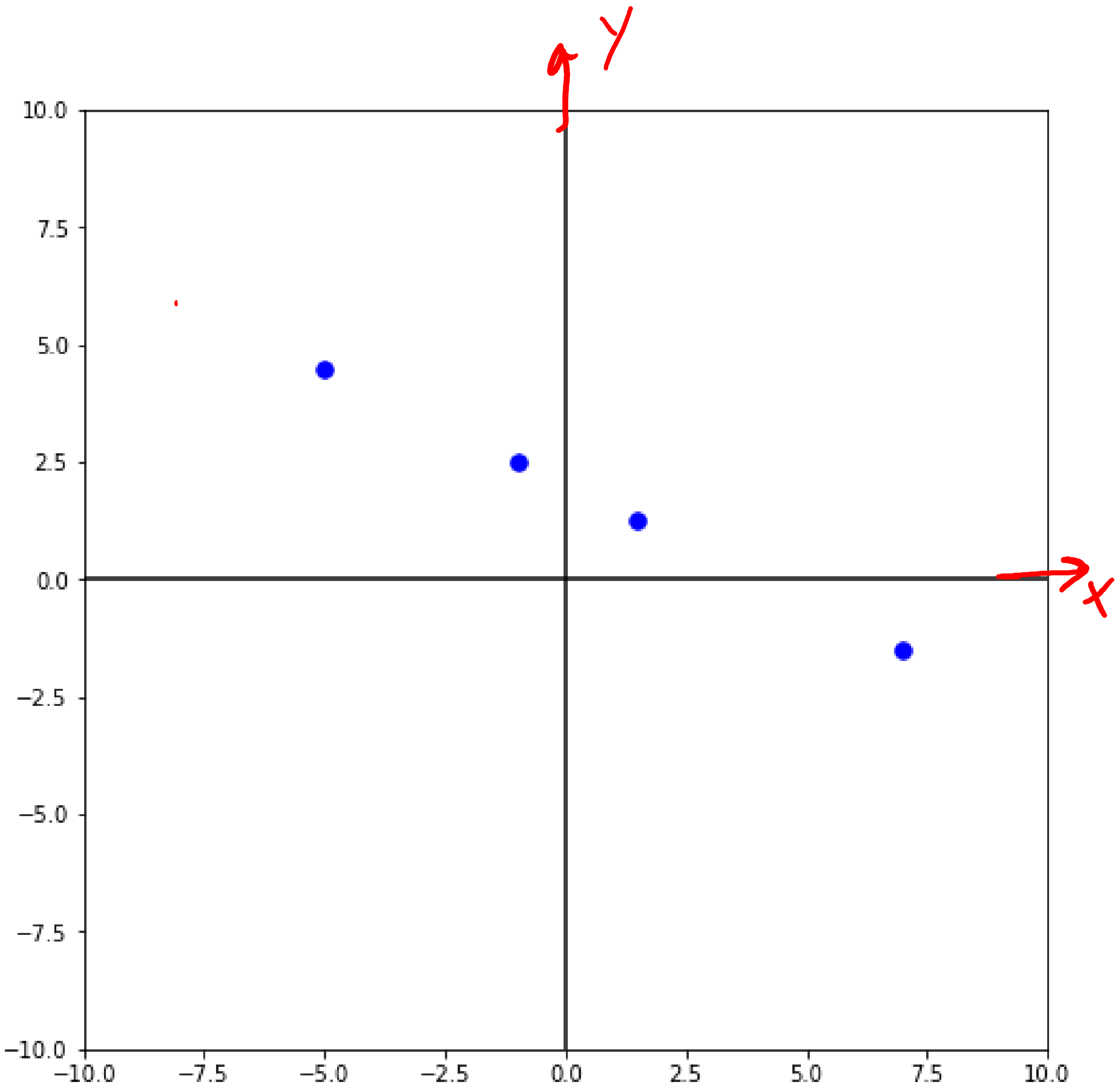
Can we fit a linear model to this data?

$$\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{4}$$
$$= \{(-1,\ 2.5),\ (7,\ -1.5),\ (-5,\ 4.5),\ (1.5,\ 1.25)\}$$

Yes 96%

No 4%

Actually, It depends on what we mean by "linear"



Linear and Affine

Notation alert!

$\forall$

$\mathbb{R}^M$

Linear combination (of a set of terms)

Multiplying each term by a scalar and adding the results

e.g. Given a set of terms $\mathcal{S} = \{x_1, x_2, x_3\}$, where $x_i \in \mathbb{R} \underbrace{\forall\ i \in \{1 \dots 3\}}$

$w_1 x_1 + w_2 x_2 + w_3 x_3$ is a linear combination of $\mathcal{S}$ if $w_i \in \mathbb{R}\ \forall\ i \in \{1 \dots 3\}$

e.g. Given a set of terms $\mathcal{S} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$, where $\boldsymbol{v}_i \in \mathbb{R}^M\ \forall\ i \in \{1 \dots 3\}$

$w_1 \boldsymbol{v}_1 + w_2 \boldsymbol{v}_2 + w_3 \boldsymbol{v}_3$ is a linear combination of $\mathcal{S}$ if $w_i \in \mathbb{R}\ \forall\ i \in \{1 \dots 3\}$

Linear and Affine

Affine combination (of a set of terms)

Affine allows for an additional scalar term to be added to a linear combination. Often called an offset or bias term

$$w_1 x_1 + w_2 x_2 + w_3 x_3 \boxed{+\, b,} \text{ where } b \in \mathbb{R}$$

$$w_1 \boldsymbol{v}_1 + w_2 \boldsymbol{v}_2 + w_3 \boldsymbol{v}_3 \boxed{+\, b,} \text{ where } b \in \mathbb{R}$$

Linear vs Affine Models

What linear usually means depends on the domain:

Linear algebra:

Linear usually means strictly linear

$y = Ax$

Geometry, algebra:

Linear usually means affine

(covered later)

$y = Ax + b$

Machine learning:

Linear usually means affine but we often transform affine to strictly linear to make the linear algebra and notation easier

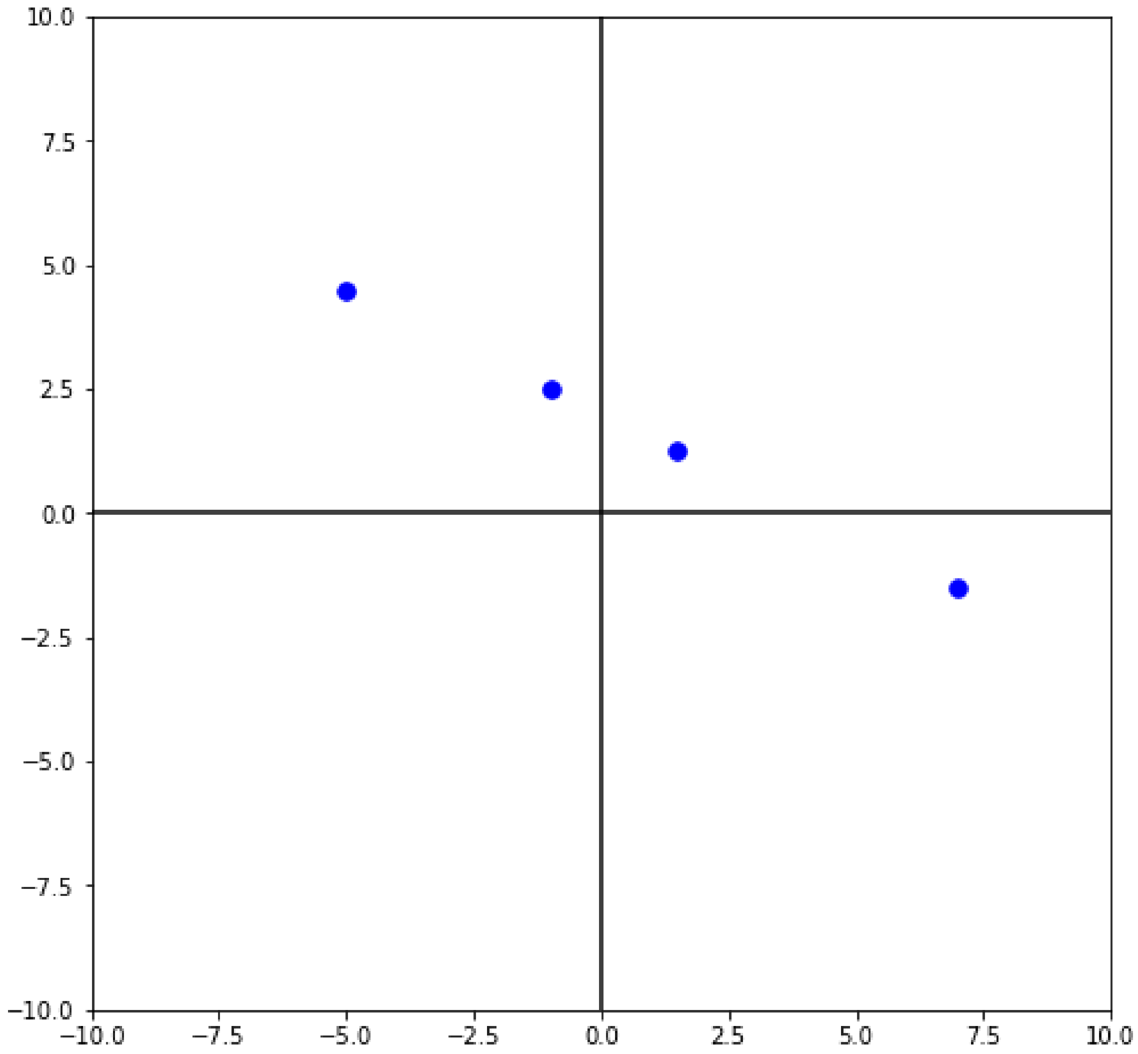
Poll 3

Can we fit a linear model to this data?

$$\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{4}$$
$$= \{(-1,\ 2.5),$$
$$(7,\ -1.5),$$
$$(-5,\ 4.5),$$
$$(1.5,\ 1.25)\}$$

$y = mx + b$

$y = mx$



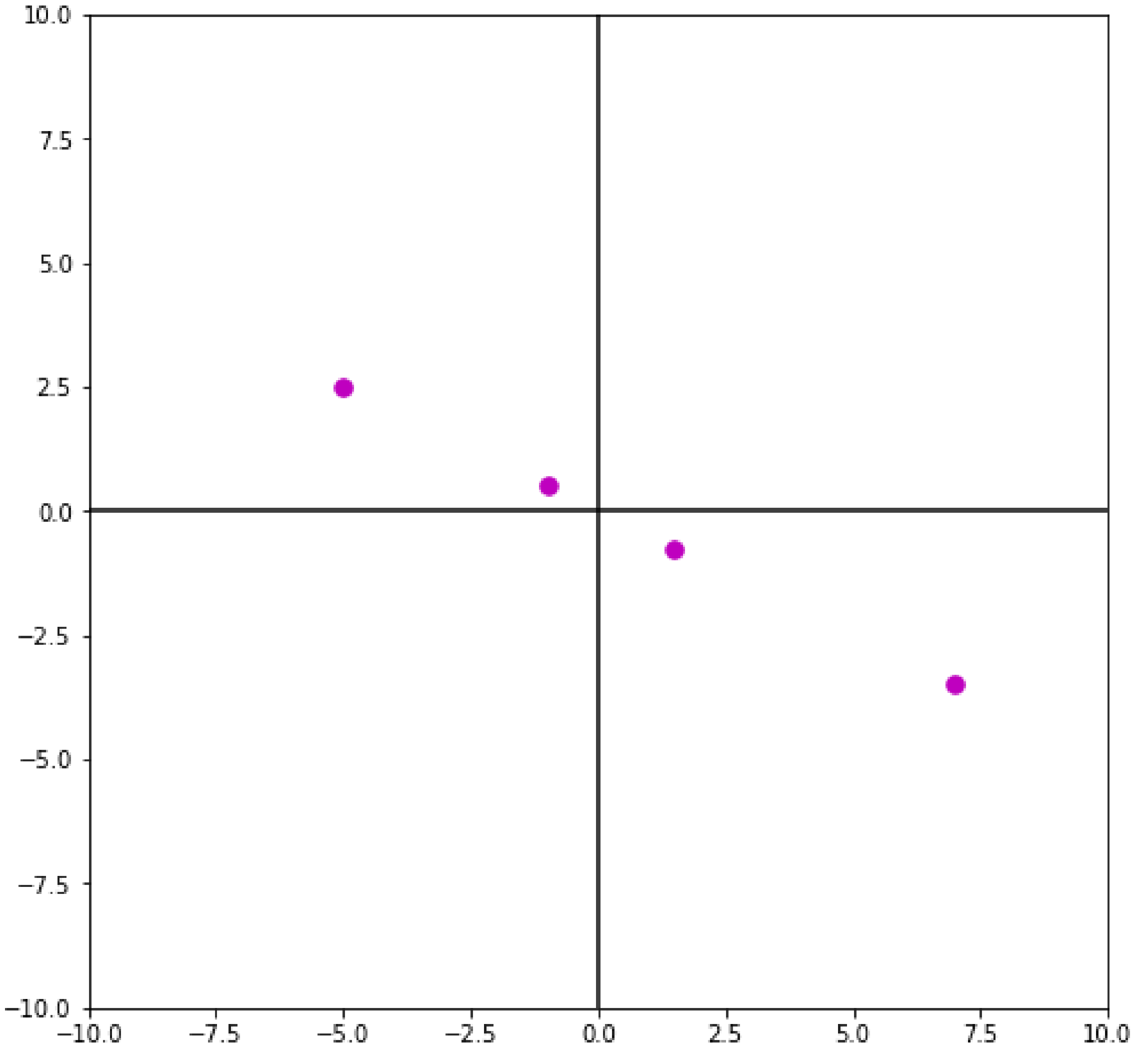
Linear (but not affine) model will include the origin

We don't need an offset (or bias) term to fit this data

$$\mathcal{D} = \left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{4}$$

$$= \{(-1,\ 0.5),$$
$$(7,\ -3.5),$$
$$(-5,\ 2.5),$$
$$(1.5,\ -0.75)\}$$

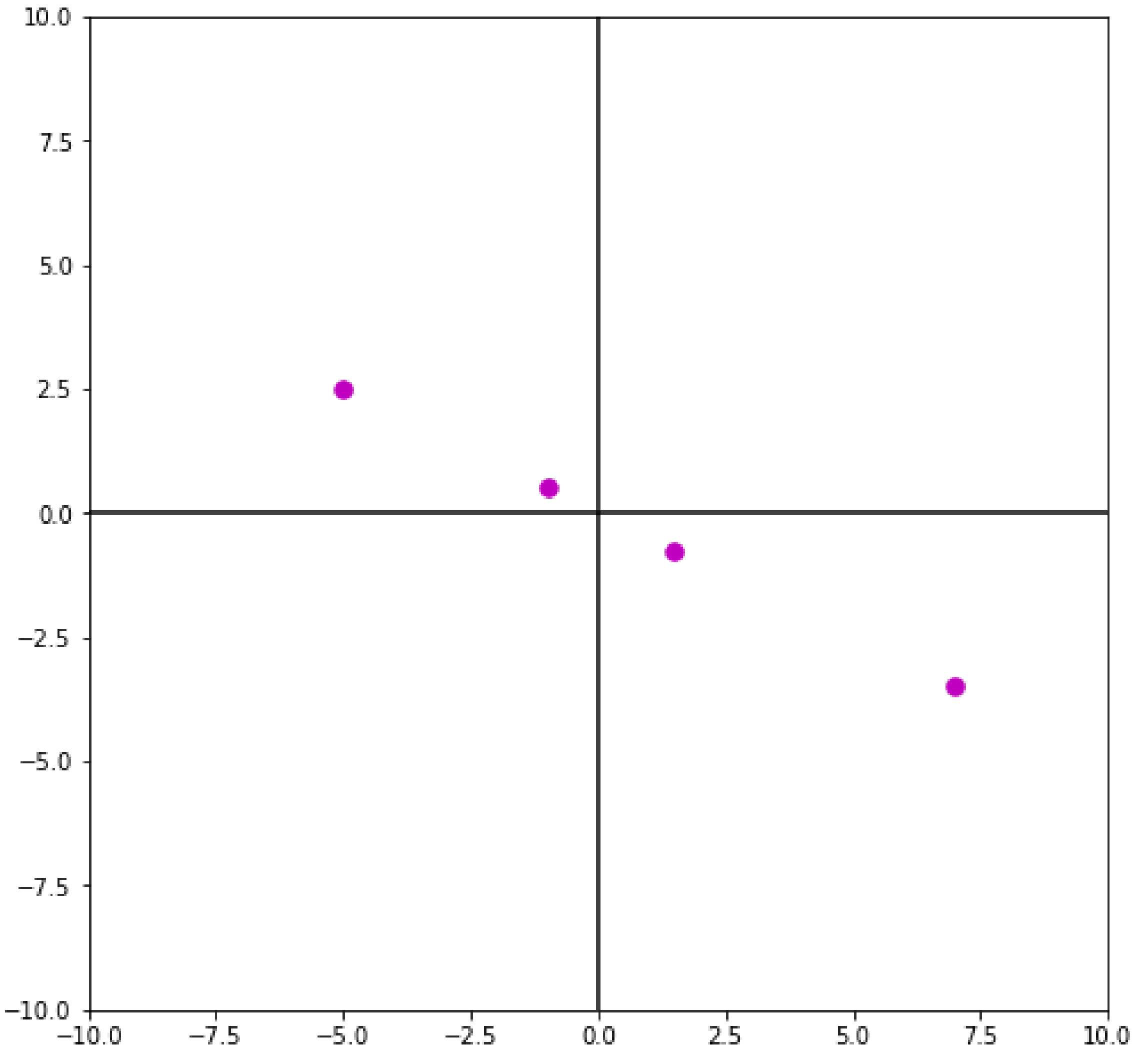
$y = mx$



Linear (but not affine) model will include the origin

We don't need an offset (or bias) term to fit this data

$$
\begin{aligned}
\mathcal{D} &= \left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{4} \\
&= \{(-1,\ 0.5), \\
&\quad (7,\ -3.5), \\
&\quad (-5,\ 2.5), \\
&\quad (1.5,\ -0.75)\}
\end{aligned}
$$

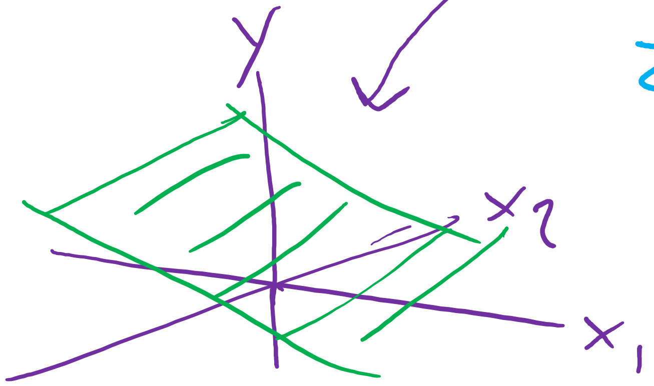


Linear (Affine) in Higher Dimensions

What are these linear shapes called for 1-D, 2-D, 3-D, M-D input?

	$\boldsymbol{x} \in \mathbb{R}$	$\boldsymbol{x} \in \mathbb{R}^2$	$\boldsymbol{x} \in \mathbb{R}^3$	$\boldsymbol{x} \in \mathbb{R}^M$
	$w \in \mathbb{R}$	$\vec{w} \in \mathbb{R}^2$		
$y = \boldsymbol{w}^T \boldsymbol{x} + b$	line	plane	hyperplane	hyperplane

$y = w_1 x_1 + w_2 x_2 + b$



$z = ax + by + c$