Announcements

HW1

- Due Mon 9/20, 11:59 pm
- Online + Written components
- Reach out for help (probability, LaTeX, anything!)

Drop deadline

- Fri 9/17

Quizzes

Vocab and Notation doc

- Notes section added!

Mathematical Foundations for Machine Learning

Linear Regression

Instructor: Pat Virtue

Today

Linear regression

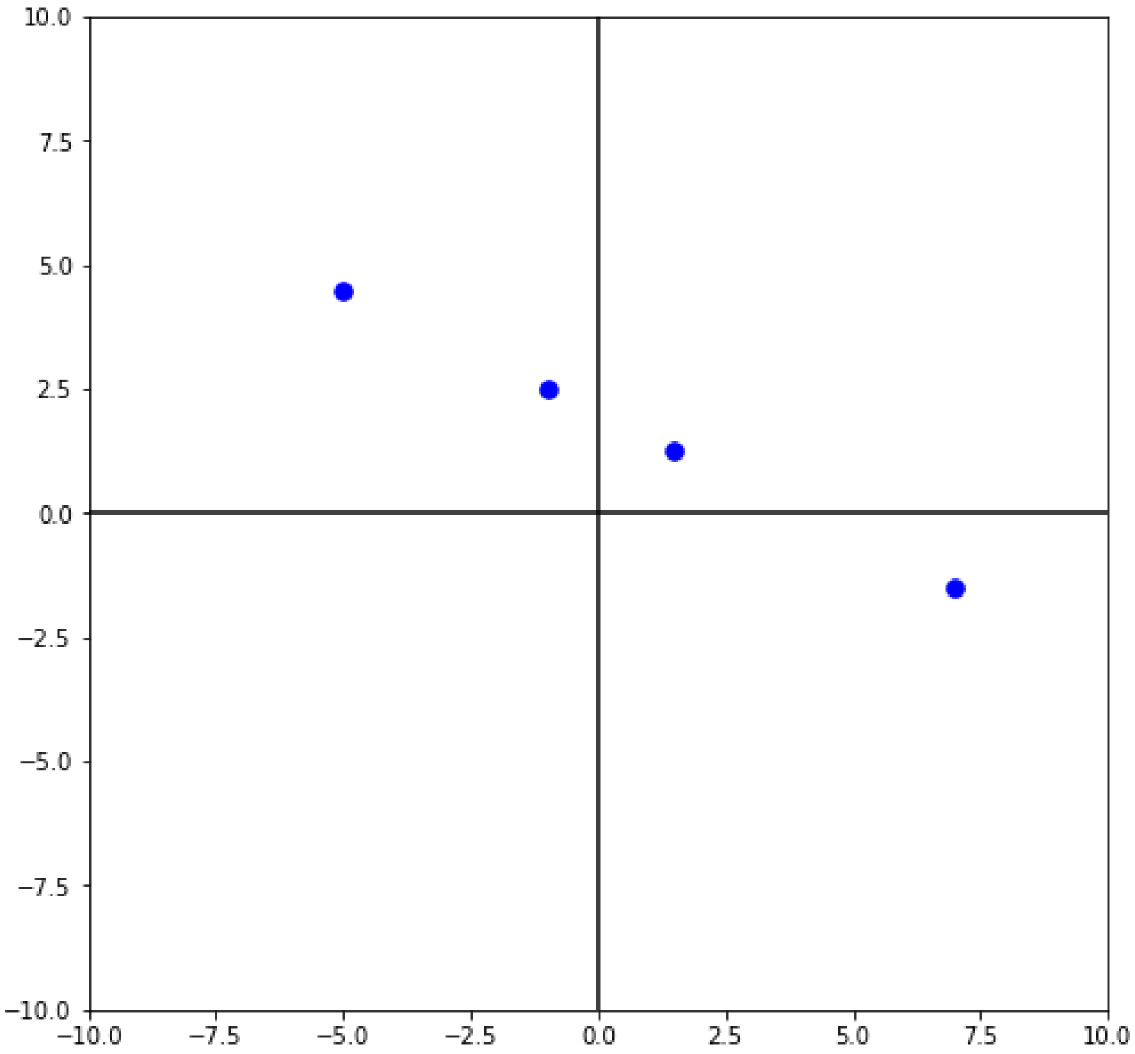
- Linear algebra formulation
  - Linear algebra properties
- Calculus
  - Multivariate calculus
  - Calculus with linear algebra
- Optimization
  - Convex functions (briefly)
  - Closed-form solutions

Linear Regression

As a reason to learn linear algebra, calculus, and optimization

Last time: Fitting a linear model

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



Last time

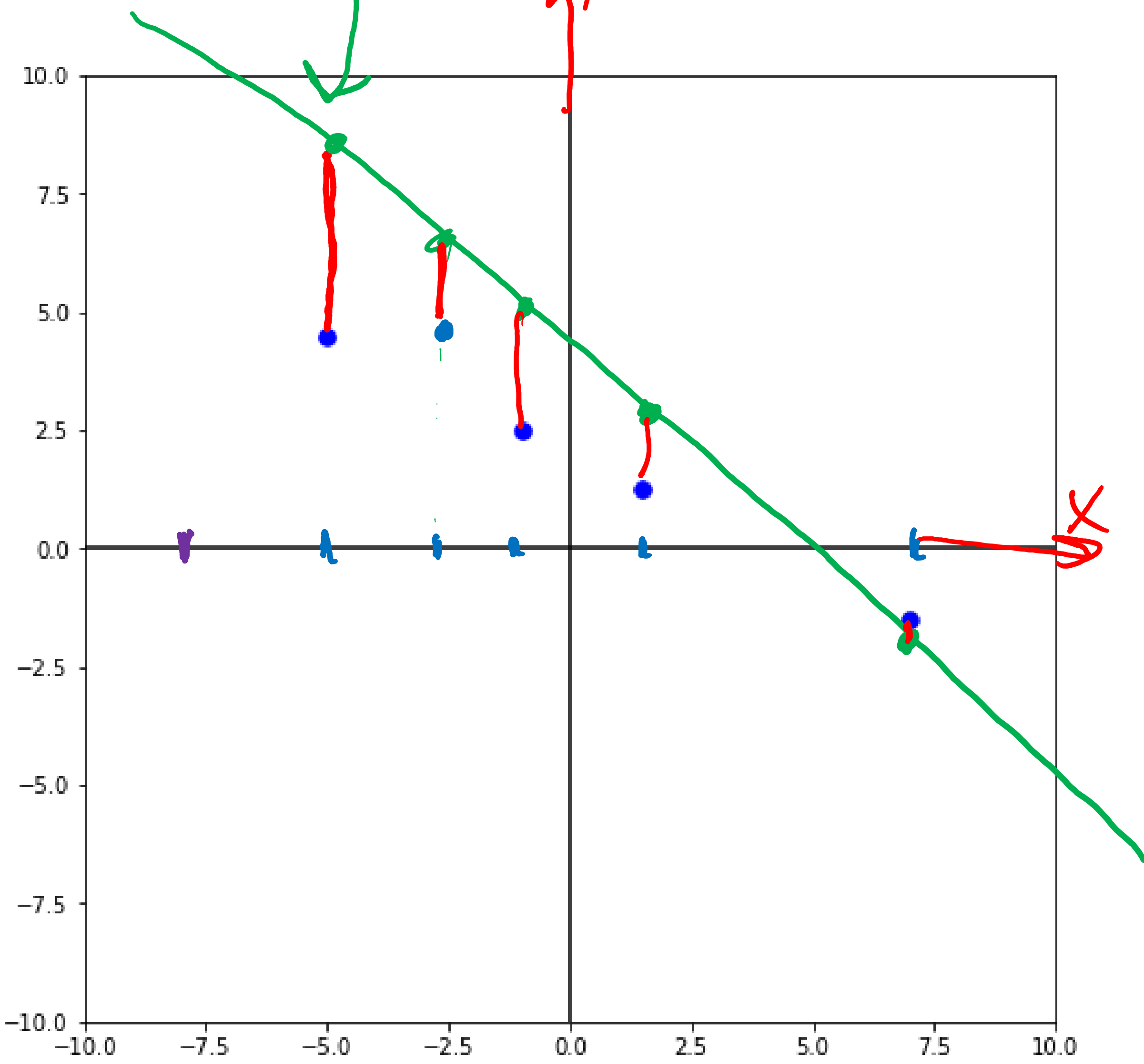
$$\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{4}$$

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

$$\sum_{i=1}^{N} \Big(\underbrace{y^{(i)} - \hat{y}^{(i)}}_{\text{error}}\Big)^2$$

$\hat{y} = mx + b$

$x_{new} = -8$



Linear Regression

Last recitation

1) Found m that minimized MSE with model $y = mx$, $x \in \mathbb{R}$
2) Found m, b that minimized MSE with model $y = mx + b$, $x \in \mathbb{R}$

Tons of summations!

Today

Linear algebra formulation

1) Find m that minimizes MSE with model $y = mx$, $x \in \mathbb{R}$
2) Find $\mathbf{w}$ that minimizes MSE with model $y = \mathbf{w}^T\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^2$
3) Find $\mathbf{w}, b$ that minimizes MSE with model $y = \mathbf{w}^T\mathbf{x} + b$, $\mathbf{x} \in \mathbb{R}^M$

Poll 1

True or False

For any matrices $A, B \in \mathbb{R}^{M \times N}$, the following always holds:

$$(A+B)^T = A^T + B^T$$

Poll 2

True or False

For any matrices $A, B \in \mathbb{R}^{M\times N}$, $C \in \mathbb{R}^{K\times M}$, the following always holds:

✗ $$CA + CB = (A+B)C$$

$[K \times M][M \times N]$

$AC + BC$

$M \times N \; K \times M$

✓ $CA + CB = C(A+B)$

$AC + BC = (A+B)C$

Poll 3

Select ALL that apply

Which of the following are equal to $\|z\|_2^2$ for any vector $z \in \mathbb{R}^N$?

A. $z^T z$ ←

B. $z z^T$

C. $\begin{bmatrix} z_1^2 \\ \vdots \\ z_N^2 \end{bmatrix}$

D. None of the above

E. I have no idea

$$\|z\|_2^2 = \sum_{i=1}^{N} z_i^2$$

Poll 4

Select ALL that apply

Which of the following hold for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ and $A, B \in \mathbb{R}^{M \times N}$?

✓ A. $\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}$

B. $A^T B \neq B^T A$ 9%

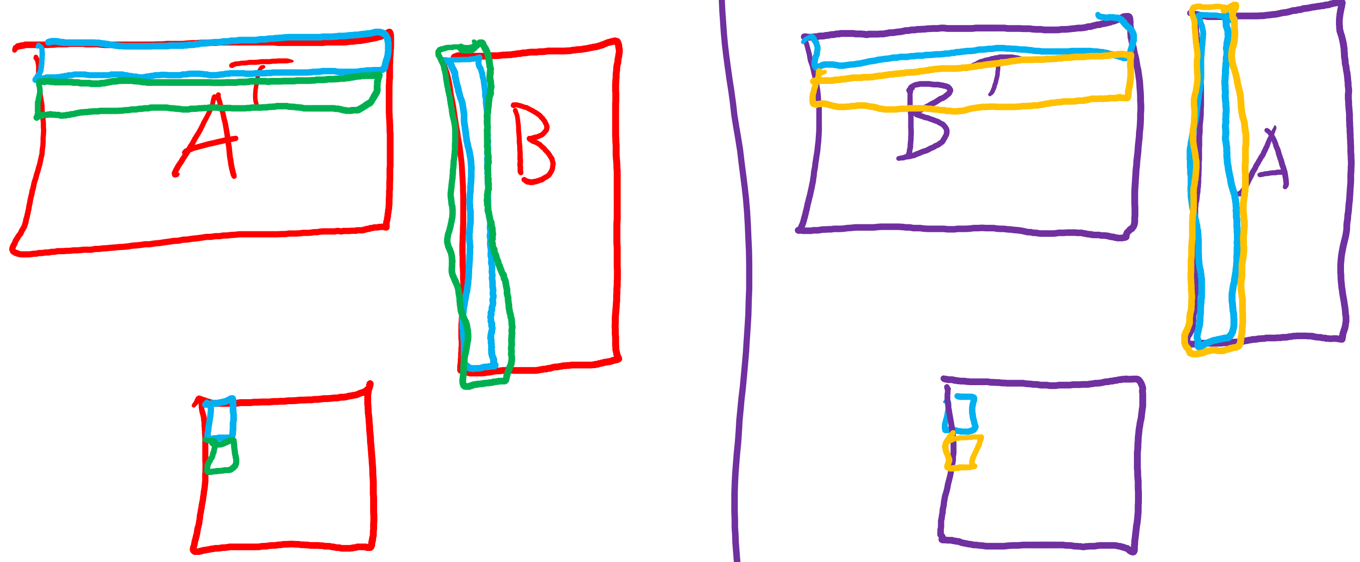
C. None of the above

D. I have no idea

$A^T B \leftrightarrow B^T$

$[N \times M]\,[M \times N]$ $\quad$ $[N \times M]\,[M \times N]$

$[N \times N]$ $\quad$ $[N \times N]$



Linear Algebra

Properties

$$(A+B)^T = (A^T + B^T)$$

$$(AB)^T \neq A^T B^T$$

The following hold for any $A, B \in \mathbb{R}^{M \times N}$

$$(ABC)^T = C^T B^T A^T$$

- ✓ $(AB)^T = B^T A^T$
- ✓ $\left(A^T B\right)^T = B^T A$
- ✓ $\left(AB^T\right)^T = B A^T$

The following hold for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ ✓

- $\mathbf{u}^T \mathbf{v} = \mathbf{v}^T \mathbf{u}$

(Why is this ok if we can't say $A^T B = B^T A$)

Linear Regression

Linear algebra formation

Find m that minimizes MSE with model $\hat{y} = mx, \quad x \in \mathbb{R}$

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

$x^{(i)}$	$y^{(i)}$
2	0
0	-1
-3	0
1	2

Linear Regression

Linear algebra formation

$\vec{y} = m\vec{x}$

Find m that minimizes MSE with model $\hat{y} = mx, \quad x \in \mathbb{R}$

$$J(m; \mathbf{x}, \mathbf{y}) = \frac{1}{N} \|\mathbf{y} - m\mathbf{x}\|_2^2$$

$$\frac{1}{N} \sum_{i=1}^{N} \left(y^{(i)} - m x^{(i)}\right)^2$$

$$= \frac{1}{N} [\mathbf{y}^T\mathbf{y} - 2m\mathbf{y}^T\mathbf{x} + m^2\mathbf{x}^T\mathbf{x}]$$

$\hat{y}^{(i)}$

$\vec{x}$ $\vec{y}$

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

$x^{(i)}$	$y^{(i)}$
2	0
0	-1
-3	0
1	2

Derive the following expansion $J(m;\mathcal{D})$

Figure out each step, given the provided justification

1 input feature
No bias term
$\hat{y} = mx \qquad m, x \in \mathbb{R}$

All data
$\mathbf{y}, \mathbf{x} \in \mathbb{R}^N$

	Justification
$J(m; \mathbf{x}, \mathbf{y}) = \frac{1}{N} \lVert \overbrace{\mathbf{y} - m\mathbf{x}}^{z} \rVert_2^2$	
$= \frac{1}{N} (y - mx)^T (y - mx)$	$\lVert \mathbf{z} \rVert_2^2 = \mathbf{z}^T \mathbf{z}$
$= \frac{1}{N} (y^T - m x^T)(y - mx)$	$(A+B)^T = A^T + B^T$
$= \frac{1}{N} [y^T y - m y^T x - m x^T y + m^2 x^T x]$	$(A+B)C = AC + BC$ and $C(A+B) = CA + CB$
$= \frac{1}{N} [\mathbf{y}^T \mathbf{y} - 2m\mathbf{y}^T \mathbf{x} + m^2 \mathbf{x}^T \mathbf{x}]$	$\mathbf{u}^T \mathbf{v} = \mathbf{v}^T \mathbf{u}$

Linear Regression

Linear algebra formation

Find m that minimizes MSE with model $\hat{y} = mx, \quad x \in \mathbb{R}$

$$J(m; \boxed{\mathbf{x}, \mathbf{y}}) = \frac{1}{N} \|\mathbf{y} - m\mathbf{x}\|_2^2$$

$$J(m) = \frac{1}{N} [\underline{\mathbf{y}^T\mathbf{y}} - 2m\underline{\mathbf{y}^T\mathbf{x}} + m^2\underline{\mathbf{x}^T\mathbf{x}}]$$

What shape is $J(m)$?

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

$x^{(i)}$	$y^{(i)}$
2	0
0	-1
-3	0
1	2

$am^2 + bm + c$

Poll 5

True or False

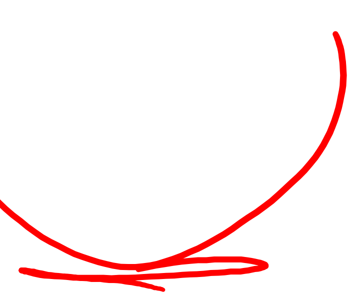
The following MSE objective function is always a parabola going up (a U shape, rather than ∩) regardless of the data in $\mathbf{x}, \mathbf{y}$:

$$J(m; \mathbf{x}, \mathbf{y}) = \frac{1}{N} \|\mathbf{y} - m\mathbf{x}\|_2^2$$

$$= \frac{1}{N} [\mathbf{y}^T\mathbf{y} - 2m\mathbf{y}^T\mathbf{x} + m^2\mathbf{x}^T\mathbf{x}]$$

Linear Regression

$$J(m; \mathbf{x}, \mathbf{y}) = \frac{1}{N}\left[\mathbf{y}^T\mathbf{y} - 2m\mathbf{y}^T\mathbf{x} + m^2\mathbf{x}^T\mathbf{x}\right]$$



$$\frac{dJ}{\partial m} = \frac{1}{N}\left[0 - 2y^Tx + 2mx^Tx\right]$$

$$0 = -y^Tx + mx^Tx$$

$$\hat{m} = \frac{y^Tx}{x^Tx} \leftarrow 0$$

Linear and Convex Functions

Optimization

Linear function

If $f(\boldsymbol{x})$ is linear, then:

- $f(\boldsymbol{x} + \boldsymbol{z}) = f(\boldsymbol{x}) + f(\boldsymbol{z})$
- $f(\alpha\boldsymbol{x}) = \alpha f(\boldsymbol{x}) \quad \forall\alpha$
- $f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{z}) = \alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{z}) \quad \forall\alpha$

Optimization

Convex function

If $f(\boldsymbol{x})$ is convex, then:

- $f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{z}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{z}) \quad \forall\, 0 \leq \alpha \leq 1$

Convex optimization

If $f(\boldsymbol{x})$ is convex, then:

- Every local minimum is also a global minimum ☺