

CS 260: Foundations of Data Science

Prof. Thao Nguyen

Fall 2024



HVERFORD
COLLEGE

Admin

- **Lab 3** grades & feedback will hopefully be posted tomorrow (Thursday)
- **Mid-semester feedback form (link on Piazza)**
- **Midterm 1 handed out today in class**
 - Do not open until you are ready to take it!
 - 3 hour time limit
 - Due Wednesday at the beginning of class (Oct 9)

Why do we have a exam?

- Process of synthesizing the material on your own is essential
- Preparing the “study sheet” is designed to facilitate that process

Outline for today

- Review
 - Linear regression
 - Gradient descent
 - Classification
 - Single feature models / decision trees
 - Evaluation metrics

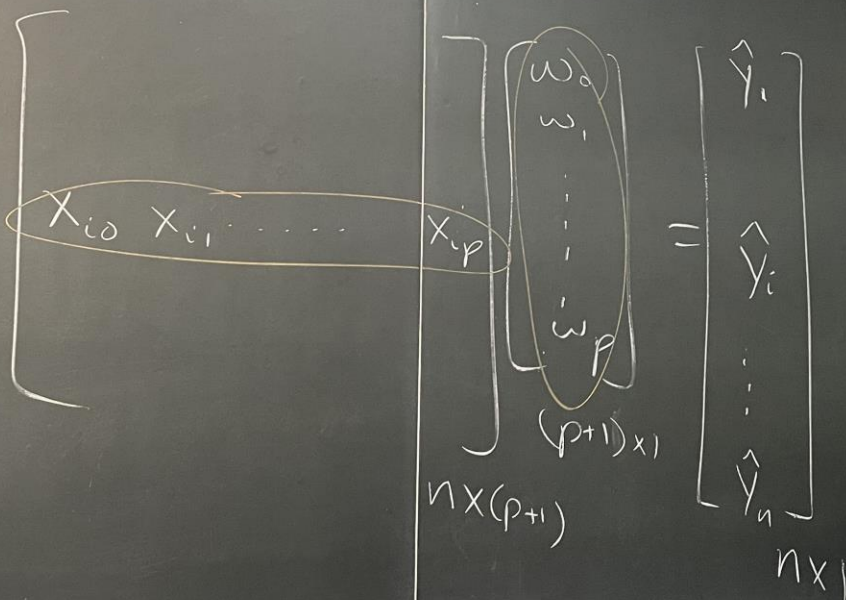
Multiple linear regression

$$) = \vec{w} \cdot \vec{x} = w_0 x_0 + w_1 x_1 + \dots + w_p x_p$$

fake!

predict(X, \vec{w})

$$X \vec{w} =$$



cost(X, \vec{y}, \vec{w})

$$J(\vec{w}) = \frac{1}{2} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

$$= \frac{1}{2} (\hat{y} - \vec{y}) \cdot (\hat{y} - \vec{y})$$

$$= \frac{1}{2} (X \vec{w} - \vec{y}) \cdot (X \vec{w} - \vec{y})$$

page 3 #7

Multiple linear regression vs. polynomial regression

multiple linear regression

$p > 1$

$$h_{\vec{w}}(\vec{x}) = w_0 + w_1 x_1 + \dots + w_p x_p$$

polynomial regression

$d = \text{deg}$

$p = 1$

$$h_{\vec{w}}(\vec{x}) = w_0 + w_1 x + w_2 x^2 + \dots + w_d x^d$$

"Simple"
linear
regression

$$X_d = \begin{bmatrix} 1 \\ x_1 \\ x_1^2 \\ x_1^3 \\ \dots \\ x_1^d \end{bmatrix} \leftarrow \vec{x}$$

Time varying η (SGD)

(learning rate) step size

$$t = 1$$

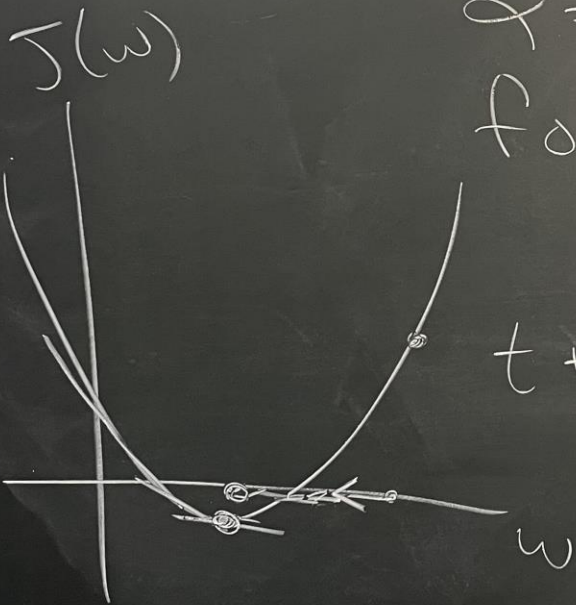
while not converged:

$$\eta = \frac{1}{t}$$

for $i = 1 \dots n$

$$\vec{w} \leftarrow \vec{w} - \eta (h_{\vec{w}}(x_i) - y_i) \vec{x}_i$$

$$t += 1$$



derivative of J wrt x_i

SGD solution to linear regression

Matrix / Vector form of SGD

while not converged:
shuffle the data.

Sort indices
4, 7, 2, 1 ... instead

for $i = 1, 2, \dots, n$:

check
cost
not
changing

$$\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$$

$$\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$$

$$- \eta (h_{\vec{w}}(\vec{x}_i) - y_i)$$

$$\begin{bmatrix} x_{i0} \\ x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix}$$

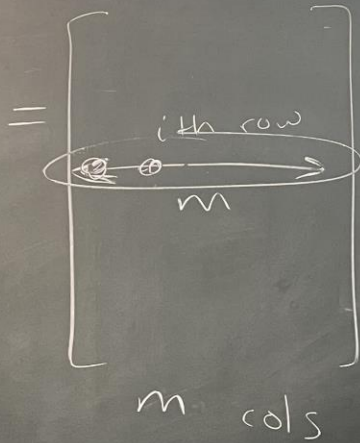
$$\rightarrow h_{\vec{w}}(\vec{x})$$

Scalar

Runtime of matrix operations

Runtime

A B
 $n \times m$ $m \times p$
 match



one entry

$$O(m) \uparrow$$

$$2m-1$$

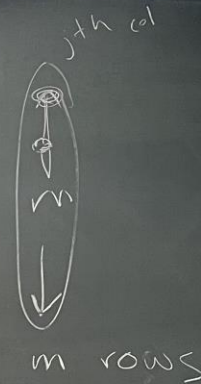
Matrix multiplication ??

$$O(npm)$$

inverse $\Rightarrow O(n^3)$

if n, p, m similar

$$\Rightarrow O(n^3)$$



$$\vec{w} = (X^T X)^{-1} X^T \vec{y}$$

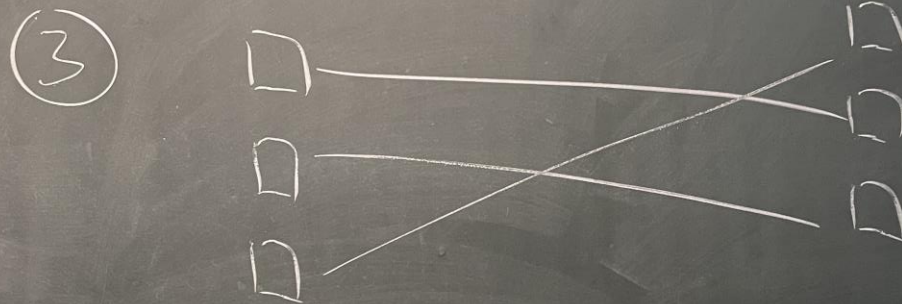
Diagram showing the dimensions of the matrices in the equation above: $(p+1) \times n$ for X^T , $n \times (p+1)$ for X , and $n \times p$ for $X^T \vec{y}$. A box labeled "(i,j)" is shown below the X matrix.

Analytic Solution
 to
linear regression

Handout 10, page 1

① classification (multi-class)

② regression

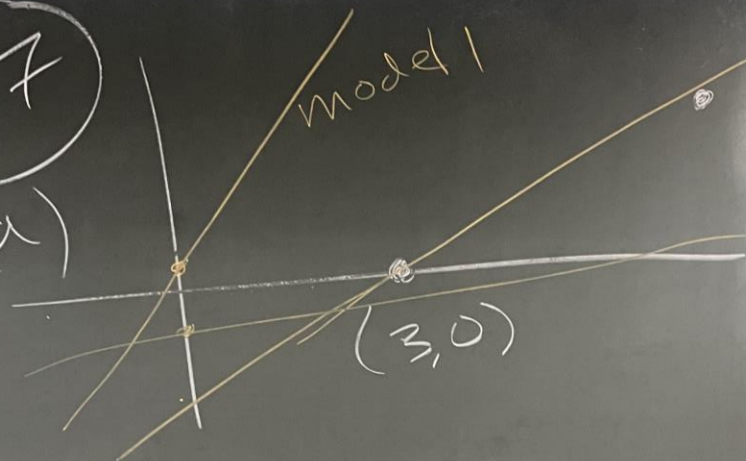


④ (a) +

Handout 10, #7

#7

(a)



$$\hat{w}_1 = \frac{1-0}{7-3} = \frac{1}{4}$$

$$y - 0 = \frac{1}{4}(x - 3)$$

$$y = -\frac{3}{4} + \frac{1}{4}x$$

w_0 w_1

(7, 1)

(b)

(c)

Handout 10, #7

(7, 1)

$$(b) \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \leftarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \eta \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 7 \end{bmatrix} - 1 \right) \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

$0.1 + 0.7 = 0$

$$\leftarrow \begin{bmatrix} 0.1 \\ 0.7 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 3 \\ 1 & 7 \end{bmatrix} \begin{matrix} \rightarrow \vec{x}_1 \\ \rightarrow \vec{x}_2 \end{matrix}$$

$$(c) \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \leftarrow \begin{bmatrix} 0.1 \\ 0.7 \end{bmatrix} - 0.1 \left(\begin{bmatrix} 0.1 \\ 0.7 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix} - 0 \right) \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$0.1 \cdot 1 + 0.7 \cdot 3 = 2.2$

$$\leftarrow \begin{bmatrix} -0.12 \\ 0.04 \end{bmatrix}$$