CS 260: Foundations of Data Science

Prof. Thao Nguyen

Fall 2024

HAVERFORD COLLEGE

Materials by Sara Mathieson

Admin

- Sit somewhere new
- GitHub Classroom submissions
- Late days
- Lab 3 will be done in pairs, please find a partner

Outline for today

- Why are models useful? (recap)
- Linear models (recap)
- Fitting a linear model (one feature)
- Model complexity and evaluation

Outline for today

- Why are models useful? (recap)
- Linear models (recap)
- Fitting a linear model (one feature)
- Model complexity and evaluation

Why are models useful?

- Understand/explain/interpret the phenomenon
- Predict outcomes for future examples

Outline for today

- Why are models useful? (recap)
- Linear models (recap)
- Fitting a linear model (one feature)
- Model complexity and evaluation

Goals of fitting a linear model

1) Which of the features/explanatory variables/predictors (x) are associated with the response variable (y)?

2) What is the relationship between x and y?

3) Can we predict y given a new x?

4) Is a linear model enough?

Linear Regression

- Output (y) is continuous, not a discrete label
- Learned model: *linear function* mapping input to output (a *weight* for each feature + *bias*)
- Goal: minimize the *RSS* (residual sum of squares) or *SSE* (sum of squared errors)

Outline for today

- Why are models useful? (recap)
- Linear models (recap)
- Fitting a linear model (one feature)
- Model complexity and evaluation

model

$$h_{\vec{w}} = w_0 + w_1 x = \hat{y}$$

$$\vec{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$\boxed{\text{Goal}}$ minimize

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

truth prediction

SSE: sum of squared errors

RSS: residual sum of squares

cost function

model

$$J(w_0, w_1) = \frac{1}{2}\sum_{i=1}^{n}\left(y_i - w_0 - w_1 x_i\right)^2$$

take derivative
& set to 0

$$\frac{\partial J}{\partial w_0} = 0, \quad \frac{\partial J}{\partial w_1} = 0$$

(a) $$\frac{\partial J}{\partial w_0} = -\sum_{i=1}^{n}\left(y_i - w_0 - w_1 x_i\right) = 0$$

$$-\frac{1}{n}\sum_{i=1}^{n} y_i + w_1 \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{n w_0}{n}$$

$\bar{x}$ avg of x's

$\bar{y}$ avg of y's

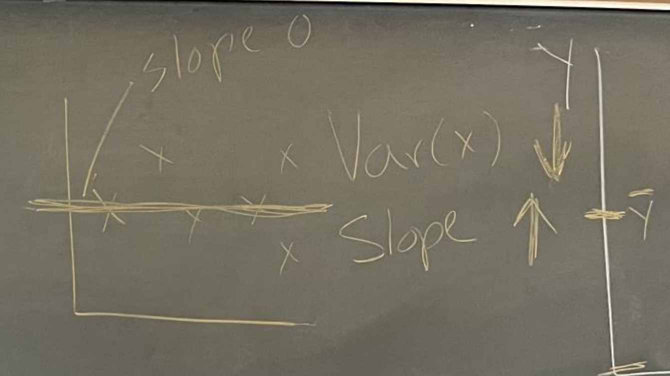
$$\boxed{\hat{w}_0 = \bar{y} - w_1 \bar{x}}$$

res

(b) $$\frac{\partial J}{\partial w_1} = -\sum_{i=1}^{n} (y_i - w_0 - w_1 x_i) x_i = 0$$

$$= -\sum_{i=1}^{n} (y_i - \bar{y} + w_1 \bar{x} - w_1 x_i) \cancel{x_i} = 0$$

$$= \sum_{i=1}^{n} (y_i x_i - \bar{y} x_i) = w_1 \sum_{i=1}^{n} (x_i^2 - \bar{x} x_i)$$



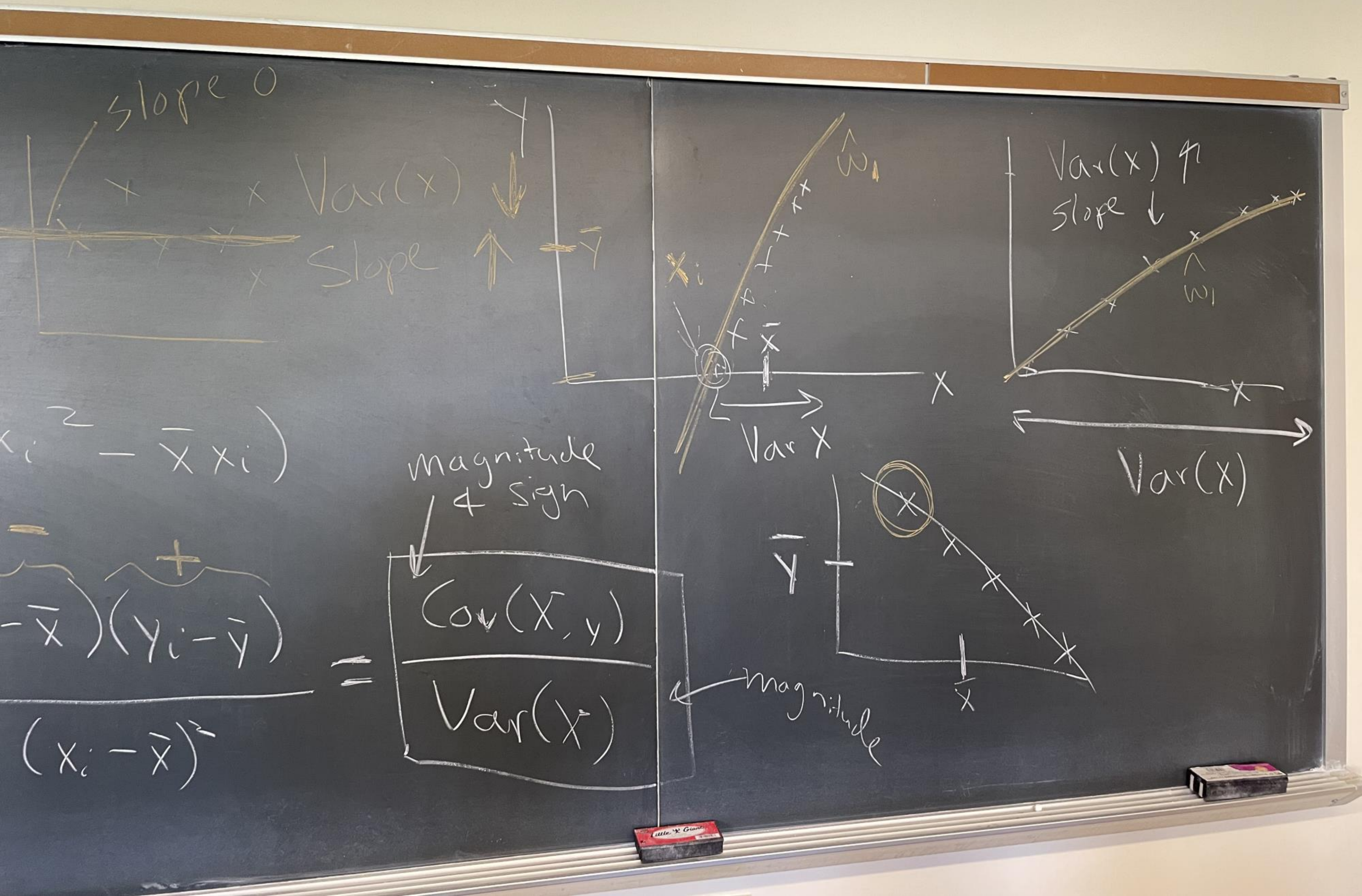
neg slope

add 0

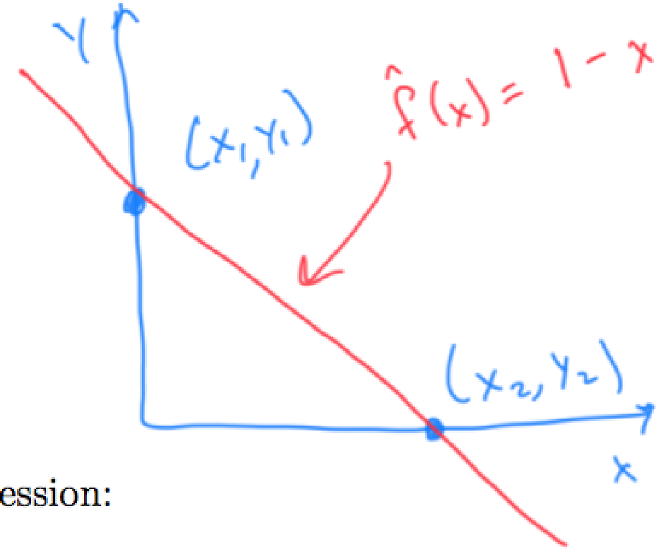
skip

magnitude & sign

$$= \hat{w}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \bar{y} x_i}{\sum_{i=1}^{n} (x_i^2 - \bar{x} x_i)} = \frac{\sum_{i=1}^{n} \underbrace{(x_i - \bar{x})}_{-} \underbrace{(y_i - \bar{y})}_{+}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \boxed{\frac{Cov(X, y)}{Var(X)}}$$


slope 0
y
Var(x)
Slope
ȳ
x_i
ŵ_1
x̄
x
Var X
Var(x) ↑
Slope ↓
ŵ_1
Var(x)
x_i² − x̄ x_i
magnitude & sign
−x̄)(y_i − ȳ)
(x_i − x̄)²
= Cov(x, y) / Var(x)
magnitude
ȳ
x̄

Handout 4 Let $n = 2$ and $p = 1$, with the following data (we will omit the first column of 1's in simple linear regression):



$$\boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

(a) Plot these two points – what should $\hat{w}_0$ and $\hat{w}_1$ be?

$\hat{w}_0 = 1$

$\hat{w}_1 = -1$

(b) This week we derived the solution for simple linear regression:

note: $\bar{x} = \frac{1}{2}$

$\bar{y} = \frac{1}{2}$

$$\hat{w}_1 = \frac{\text{Cov}(\boldsymbol{x}, \boldsymbol{y})}{\text{Var}(\boldsymbol{x})} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \hat{w}_0 = \bar{y} - \hat{w}_1 \bar{x}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Use these equations to compute $\hat{w}_0$ and $\hat{w}_1$ and verify your answer to (a).

$$\hat{w}_1 = \frac{\cancel{\tfrac{1}{2}}\left[(1-\frac{1}{2})(0-\frac{1}{2}) + (0-\frac{1}{2})(1-\frac{1}{2})\right]}{\cancel{\tfrac{1}{2}}\left[(1-\frac{1}{2})^2 + (0-\frac{1}{2})^2\right]}$$

$$= \frac{-\frac{1}{4} - \frac{1}{4}}{\frac{1}{4} + \frac{1}{4}} \Rightarrow \boxed{\hat{w}_1 = -1}$$

$$\Rightarrow \hat{w}_0 = \frac{1}{2} - (-1)\frac{1}{2}$$

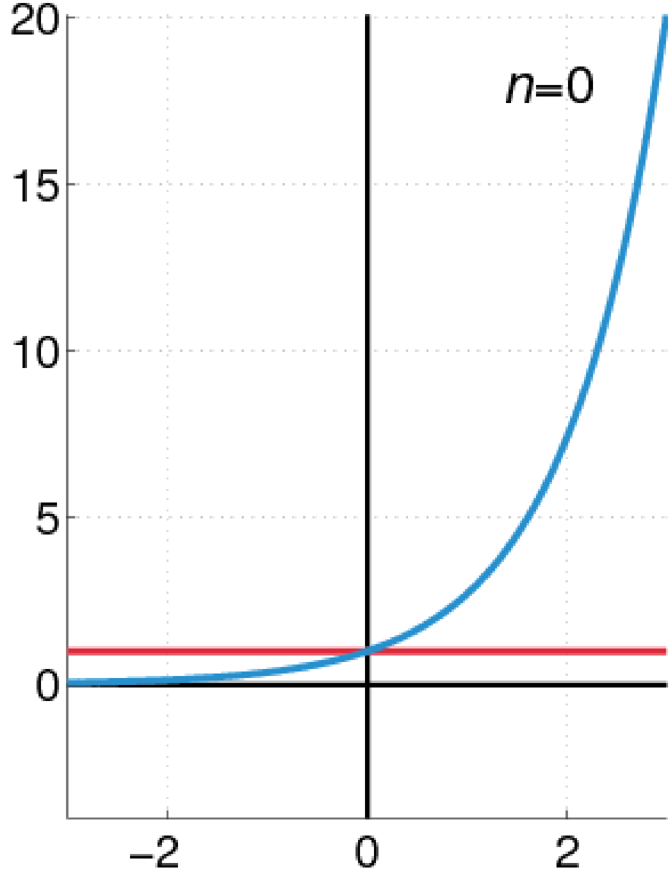
$$\Rightarrow \boxed{\hat{w}_0 = 1}$$

Outline for today

- Why are models useful? (recap)
- Linear models (recap)
- Fitting a linear model (one feature)
- Model complexity and evaluation

Model Complexity

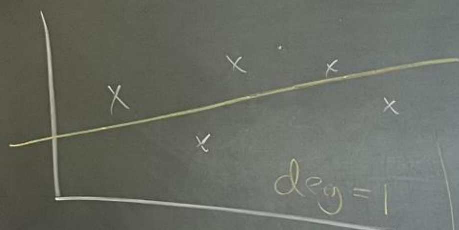
Why stop at a linear model?

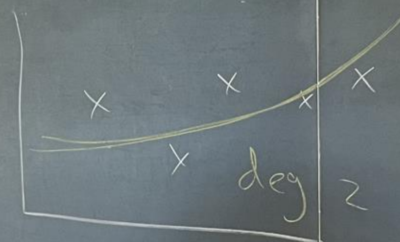


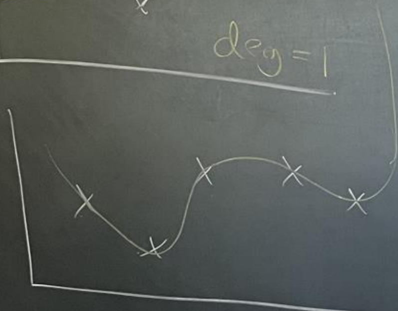
Wikipedia

Model complexity

→ why stop at linear?







$\boxed{\text{n points}}$

n-1 degree will have

$J = 0$

Elbow Plot