DATASCI 207

Nedelina Teneva, PhD

School of Information, UC Berkeley

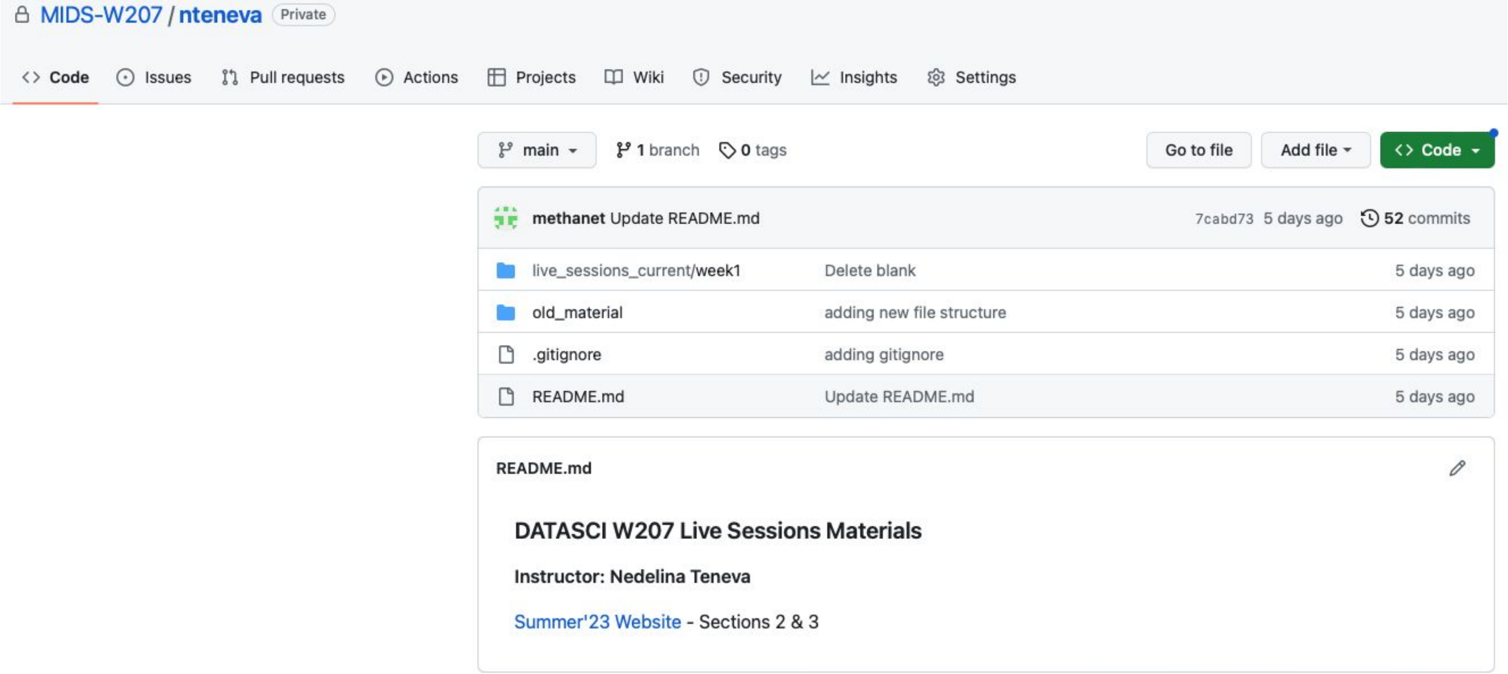
Bio

- Currently at Megagon Labs (R&D lab focusing on fundamental ML research)
- Previously at ML Science Manager at Amazon Alexa
- PhD from University of Chicago (focusing on optimization）
- Background in Molecular Biology

You?

- (Under) graduate major
- Current job/occupation (if any)
- Why a Masters in Data Science?

Announcements

- Course Website
- 

methanet Update README.md 7cabd73 5 days ago 52 commits

live_sessions_current/week1	Delete blank	5 days ago
old_material	adding new file structure	5 days ago
.gitignore	adding gitignore	5 days ago
README.md	Update README.md	5 days ago

README.md

DATASCI W207 Live Sessions Materials

Instructor: Nedelina Teneva

Summer'23 Website - Sections 2 & 3

Objectives

- Intro to our first ML technique: linear regression (LR)
- Learning about how to solve LR
- Basic LR and a Tensorflow example: review after class!
    - https://github.com/MIDS-W207/nteneva/blob/main/live_sessions_current/week2/Week_2_Linear_Regression_I.ipynb

Linear Regression

- Why do we use linear regression?
- What assumptions does LR make for the relationship between outcomes (**y**) and features (**X**)?

Linear Regression

- Why do we use linear regression?
- What assumptions does LR make for the relationship between outcome (**y**) and features (**X**)?
- **Linear Algebra notation**

X: matrix of size (n, m) - inputs/features/covariates/independent var's…

y: vector of size n (column by definition) - output/dependent var's/response…

Model: $y = X\beta + e$

β: vector of size m - parameters/weights

e: vector of size n - error/noise

Goal: estimate β s.t. the noise/error e is minimized

Linear Regression

Source: ESL II

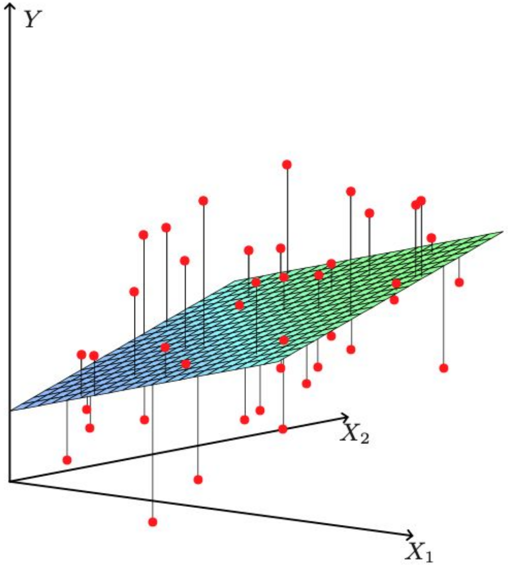


FIGURE 3.1. *Linear least squares fitting with $X \in \mathbb{R}^2$. We seek the linear function of X that minimizes the sum of squared residuals from Y.*

Example

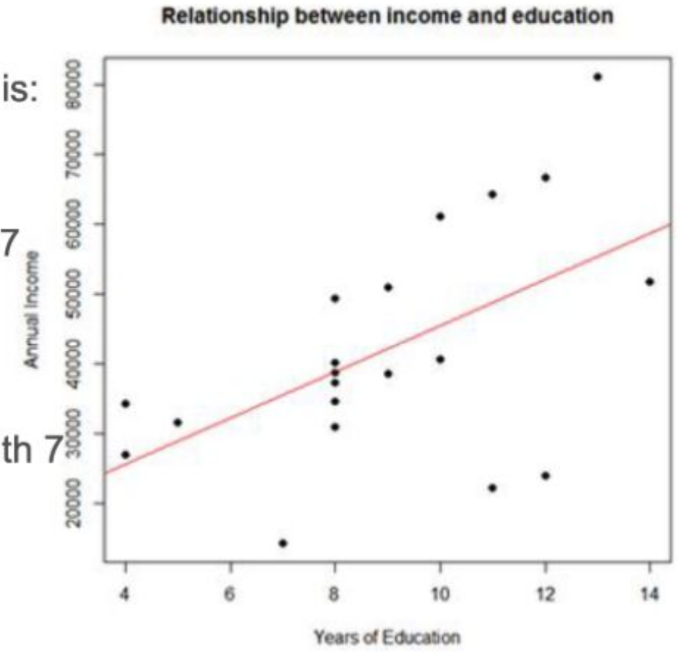
Based on the data and regression line, what is:

- the actual income for the individual with 7 years of education?

$5000

- the predicted income for an individual with 7 years of education?

$35000



How do we compute LR?

- Direct methods
    - See https://www.cs.toronto.edu/~rgrosse/courses/csc321_2017/readings/L02%20Linear%20Regression.pdf
    -
- Iteratively

What is gradient descent?

Gradient descent: key components

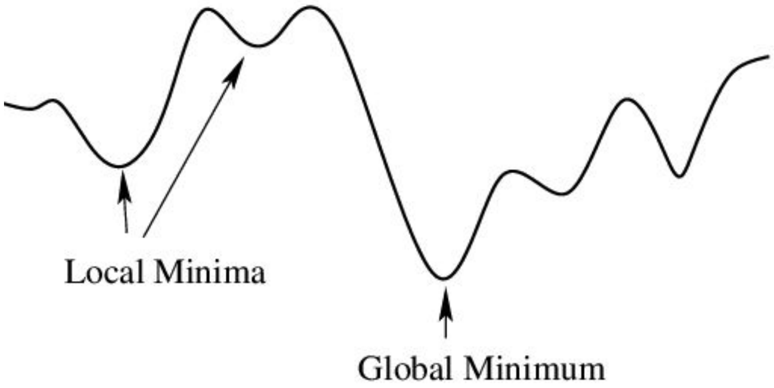
- Model
- Parameters
- Cost function
- Objective: minimize the cost function

Optimization

- What is global minim**um**?
- What is local minimum?
- How how we avoid local minim**a**?

Optimization

- What is global minim**um**?
- What is local minimum?
- How how we avoid local minim**a**?



When do we stop iterating?

When do we stop iterating?

- When the validation error stops improving (i.e., difference between step t and t +1 is below some threshold)
- Based on the loss

Hyperparameters

- What are hyperparameter? Examples?
- How are they different from the parameters?

Hyperparameters

- Examples
  - Batch size
  - Learning rate
  - Epochs

- How do we set their values?

Hyperparameters

- Examples
  - Batch size
  - Learning rate
  - Epochs

- How do we set their values?
  - Hyper parameter optimization using e.g. cross validation