Data 100/200 Homework 8 Written

Jackie Hu

TOTAL POINTS

7 / 9

QUESTION 1

1 Question 1 1 / 1

✓ **+ 1 pts** Correct

+ 0 pts Blank or incorrect

QUESTION 2

2 Question 2 1 / 1

✓ **+ 1 pts** Addresses power redistribution between all 3 groups

+ 0.5 pts Addresses power redistribution between only one or two groups

+ 0 pts Incorrect/Blank

QUESTION 3

3 Question 3 1 / 1

✓ **+ 1 pts** Answers both parts of question thoughtfully.

+ 0.5 pts Answers only one part or gives insufficient explanation

+ 0 pts Completely incorrect, off-topic, or blank

QUESTION 4

4 Question 4 1 / 1

✓ **+ 1 pts** Correct

+ 0.5 pts Identifies an aspect of inaccessibility of the current CCAO's policy.

+ 0.5 pts Suggests an improvement for the transparency implementation of CCAO

+ 0 pts Incorrect/Blank

QUESTION 5

5 Question 5 1 / 1

✓ **+ 1 pts** Thoughtful explanation naming specific aspects of CCAO's open data initiative

+ 0.5 pts Incomplete or vague explanation

+ 0 pts Incorrect, off-topic, or blank

QUESTION 6

6 Question 7a 0 / 2

+ 2 pts Fully correct loss function and partial derivatives of R with respect to $$\theta_1$$ and $$\theta_2$$

+ 1 pts Correct loss function $$R(x,y,\theta_1,\theta_2$$)

+ 0.5 pts Correct partial derivative of R with respect to $$\theta_1$$

+ 0.5 pts Correct partial derivative of R with respect to $$\theta_2$$

✓ **+ 0 pts** Incorrect/Blank

QUESTION 7

7 Question 8d 2 / 2

✓ **+ 2 pts** Correctly describes path that $$\theta$$ takes and why SGD is preferred over batch GD

+ 1 pts Missing description of path or why SGD is preferred over batch GD

+ 0 pts Incorrect/Blank

ılı gradescope

0.0.1 Question 1

In the context of estimating the value of houses, what does error mean for an individual homeowner? How does it affect them in terms of property taxes?

For individual homeowner, we need to consider the individuasl's context, the error could mean the appraisal system is unfair towards the individual's property, the information of the individual's house is wrong.

If the property has biased, the property tax is biased towards low-middle income households, which put them at a higher risk and a more vonlunarible position.

0.0.2 Question 2

How do the CCAO's transparency initiatives aim to redistribute power between the tax lawyer industry, the CCAO, and the constituents of Cook County?

Hint: You may find the following official post from CCAO helpful.

The County aimed to provide more contextm of their data, and provide the open source and options for the public to verify and appeal it's data analysis process.

They also public the modeling data sources and the assesment data source for individual to explore.

Taxpayers may then appeal to the Board of Review, after which the Board certifies the roll, and tax bills are sent to taxpayers.

The deligation of power over these housing data and assesment model could help the public to get a better understanding of the appraisal process, as well as gaining power in supervising the process.

0.0.3 Question 3

Take a look at the Residential Automated Valuation Model files under the Models subgroup in the CCAO's GitLab. Without directly looking at any code, do you feel that the documentation sufficiently explains how the residential valuation model works? Which part(s) of the documentation might be difficult for nontechnical audiences to understand?

I think the documentation is well explained, and friendly to non-tech audience. It's quite dense, but the information are indexed in a way that's easy to reference.

They also added table and graphs to explain how the data is collected, how the features are selected and how does the referencing works in model data, and how does the assesment data is used to predict house value.

0.0.4 Question 4

How could the CCAO improve its implementation of transparency? What aspects of it are inaccessible and to whom? Consider the concepts of expertise and power.

For individuals who do not have internet/computer/software access, some of the information is not accessable.

I think other than improving the transparency of the actual process, it's helpful to let public know that this resource is available for individuals to access. And maybe some sort of community support can be helpful for the individual who don't have technology access and technical background in understanding the process.

0.0.5 Question 5

How does the CCAO's open data initiative support it in maintaining people's trust in its housing assessments?

The CCAO open data initiative support helps to answer the following questions: * How do we get the data? By public the open data source individuals can see how the data/features are selected, how the process is implemented * Who is involved, who're the stakeholdes, how can individual involved in the process? It granted the access and use of the data for people who is the property holder, journalist, data scientist, social workers, education researchers to implement and oversee the process, which is transparent and helpt to eliminate potential bias in the system. * Consequences Individuals can also access, leanr and explore the appraising process if they want to.

Question 7a Recall the optimal value of θ minimizes our loss function. One way of solving for θ is by taking the derivative of our loss function with respect to θ, like we did in HW 6.

Write/derive the expressions for following values and write them with LaTeX in the space below.

- $R(\mathbf{x}, \mathbf{y}, \theta_1, \theta_2)$: our loss function, the empirical risk/mean squared error
- $\frac{\partial R}{\partial \theta_1}$: the partial derivative of R with respect to θ_1
- $\frac{\partial R}{\partial \theta_2}$: the partial derivative of R with respect to θ_2

Recall that $R(\mathbf{x}, \mathbf{y}, \theta_1, \theta_2) = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2$

- $\frac{\partial R}{\partial \theta_1}$: $2(y - \theta_1 x - sin(\theta_2 x)) - x$
- $\frac{\partial R}{\partial \theta_2}$: $2(y - \theta_1 x - sin(\theta_2 x)cos(\theta_2 x))x$

Question 8d: Analyzing Learning Rates In 1-2 sentences, describe what you notice about the path that θ takes with a static learning rate vs. a decaying learning rate and for batch vs. stochastic gradient descent. Based on these comparisons, why do we often prefer SGD over batch GD in practice? In your answer, refer to the plots above.

- Static learning rate VS. Decaying learing rate Based on the loss visualization, the decaying learning learn faster, (find the optimum step size faster); therefore could be computational more efficient, and prevent overshoot step size, since it's decreasing the step once it get closer to the minimum.

- Batch VS. stochastic Based on the loss visualization, the stochastic gradient descent takes more random small steps, they take almost the same amount of time to converge, however the intuition behind mini batch (stochastic) approach is more computational efficient and could be more useful on a larger dataset.