

BA865: Individual Homework Assignment

Submit this file, edited to include your answers to conceptual questions, along with your Collab notebook (i.e., an .ipynb file) via Blackboard. Leave rubric blank (this is for us to use in grading).

| Criteria | Comments | Grade |
|--|----------|------------|
| Comments / Formatting (3) <ul style="list-style-type: none">• Clean code, intuitive variable naming conventions, code-re-use, etc.• Comments to clarify what purpose your code blocks serve.• Clean notebook formatting, with subsections, e.g., <code>##**TEXT**</code> headers, sub-headers. | | |
| Code & Conceptual Implementation (7) <ul style="list-style-type: none">• Code executes as is without any errors.• Data imported and pre-processed.• Some assessment of relevant descriptive statistics.• Implements a NN architecture (employing appropriate layers, activations, etc.)• Fits the model using an appropriate loss function.• Implements out of sample model evaluation.• Make suggestions about things to explore going forward that may improve performance. | | |
| Conceptual Questions (5) <ul style="list-style-type: none">• Responses to conceptual questions (detailed below) are clear, coherent, and correct. | | |
| Total | | /15 |

Context & Instructions

You are a data scientist working for Boston Blue Bikes. As part of an operational improvement initiative, you have been asked to develop a predictive model that can accurately forecast, at the outset of a particular bike rental, how long the customer's rental will last (i.e., the rental / trip duration). Your predictive model will be used to inform logistics within the bike network, e.g., the reallocation of bikes across / between rental stations, to ensure bike availability. Your manager has provided you with a sample of historical data on bike rentals, including some customer details, geographic information, information on rental timing, and the rental duration. However, your manager was careful to point out that you are welcome to merge in additional predictors / features from external public sources if it helps improve your model.

Conceptual Questions

1. Draw a visual representation of the neural network that you just implemented (i.e., like the diagrams I have shown you in class, depicting the input layer, hidden layers, and the output layer). State how many weights and how many bias parameters your network includes.

The easiest way to do this is using the `plot_model()` and `model.summary()` commands. You could just paste the output here from each command.

- 2.** What loss function did you use with your neural network? Why did you choose this loss function? What activation function did you use in the output layer? Why did you choose this activation function?

MAE or MSE probably make the most sense. Binary loss or categorical cross entropy loss will obviously be wrong here. Output should probably be linear, or maybe relu, to accommodate continuous positive output.

- 3.** Choose any other activation function (one you did not use) and explain why that activation function would have been a bad choice in this setting.

Sigmoid wouldn't be capable of predicting values > 1 .

- 4.** Imagine your prediction task has changed, and your objective is not to predict trip duration, but is instead to predict trip destination (i.e., the station at which the customer is most likely to return the bike). What activation function and loss function would you use for that alternative prediction task?

Softmax with (sparse) categorical cross-entropy.

- 5.** Did you consider using any external datasets to expand your predictor set? If so, which ones? If not, why not? Did you use all the provided predictors as inputs to your neural network? If not, why not?

External data on geography might help, most obviously. Open street maps travel time between lats and longs could work. Topographical information maybe could also help.

Using stop-time is not appropriate because you won't have that feature at the time a prediction is made!

One-hot encodings of bike IDs or station IDs is something to be careful with. Your model will not be usable if a new bike ID appears, so you would have think about some error handling for that situation, e.g., a catch-all category for "new bike."

Peer Support

Indicate here what other students you spoke to when you were working on this project, to brainstorm solutions or work through problems:

ChatGPT (Bing AI, or Bard, etc.)

Document your use of any AI large-language models (LLMs) here that you used when working on the assignment. Provide the exact prompts you provided to the LLM as input. You can also provide screenshots.