Week 4 Quiz

Content: Chapter 1 “Introduction to Machine Learning with Python”

1. Name one of the machine learning example problems (book mentions “Problems Machine

Learning Can Solve”).

“Identifying the zip code from handwritten digits on an envelope” is one of the machine-learning example problems mentioned in the book.

2. What does the book say is the reason for using python when doing machine learning?

According to the book, python is popular for machine learning due to its simplicity, extensive library, large community support, and open source, which makes it easier for all levels of Python users to develop models and focus on solving problems without worrying about the underlying implementation. (I love the abstraction that python provides to the user)

3. Why do we split data into training and testing sets?

Data is split into training and testing sets to ensure that the model can generalize well to unseen data. The training set is used to learn patterns, while the testing set is used to evaluate the model’s prediction/performance, ensuring that it does not just memorize the training data but can handle new inputs effectively.

4. What is matplotlib?

Matplotlib is a visualization library in Python that is used to create static, interactive, and animated plots. It helps in visually understanding data patterns and results, which is important in machine learning for analyzing and presenting findings.

5. Explain how the nearest-neighbor algorithms discussed last week in class determine distance.

nearest-neighbor algorithms determine distance using metrics such as Euclidean distance. This involves calculating the straight-line distance between two data points in feature space. Other distance metrics like Manhattan distance can also be used depending on the nature of the problem and data.