A Multilayer Perceptron with Wine Quality Data

Name : Bhanu Prakash Mallam
ID : 24095576

Abstract

This tutorial investigates how the width of the hidden layer in a Multilayer Perceptron (MLP) affects the ability to predict wine quality from physicochemical measurements. Using the WineQT dataset, we evaluate model performance for different network widths and demonstrate the trade-off between underfitting and overfitting. The goal of this tutorial is to help readers understand how neural networks behave when model capacity changes and how to make informed choices when designing MLP architectures in practical data-science tasks.

---

1. Introduction

Modern machine-learning techniques can automate quality assessment processes that traditionally rely on expert judgement. One well-studied example is wine quality prediction, where chemical properties are used to estimate human-assigned quality ratings.

Artificial Neural Networks — specifically **Multilayer Perceptrons (MLPs)** — can learn non-linear relationships between features and output labels. However, their performance strongly depends on architectural choices, such as the number of neurons in the hidden layer.

This tutorial focuses on the following question:

How does hidden layer width influence the performance of an MLP when predicting wine quality?

We provide an end-to-end guide including dataset exploration, model implementation, experimentation, and interpretation.

---

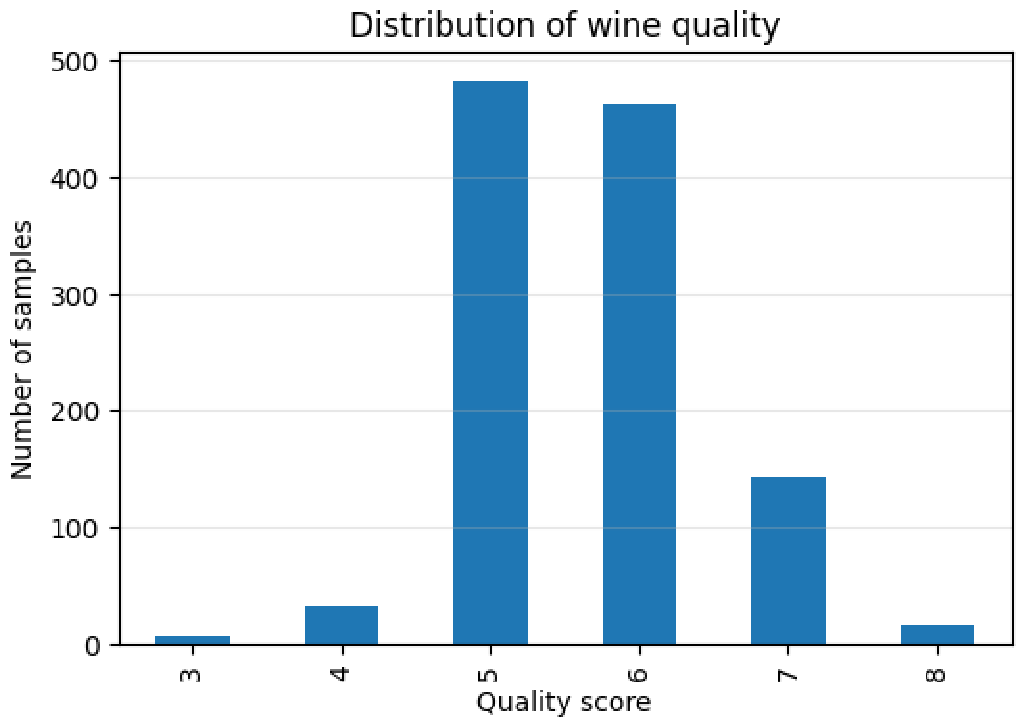
2. Dataset Overview

We use the **WineQT dataset**, a variant of the Vinho Verde wine dataset introduced by Cortez et al. (2009). Each example contains 11 physicochemical features (e.g., acidity, pH, alcohol level) and a discrete quality score typically ranging from 3–8.

Wine Quality Distribution

The dataset shows clear class imbalance: most samples are rated 5 or 6, with few at the extremes.

📊 *Figure 1 — Wine Quality Distribution*



Imbalanced data can lead to biased classifiers; therefore, performance must be judged carefully across classes.

---

3. Methodology

3.1 Data Preparation

- Removed identifier column (Id)
- Train/test split: **80% training / 20% testing**, stratified by quality scores
- Features were **standard-scaled** to improve neural-network training stability

3.2 Baseline Model

A **multinomial Logistic Regression** model serves as the baseline.
It provides a linear decision boundary, useful for comparison against the MLP.

3.3 Multilayer Perceptron Setup

- Single hidden layer

- ReLU activation
- Adam optimiser
- Same hyperparameters across experiments except hidden-layer width

Tested widths:

8, 16, 32, 64, 128 neurons

For each configuration, we measure:

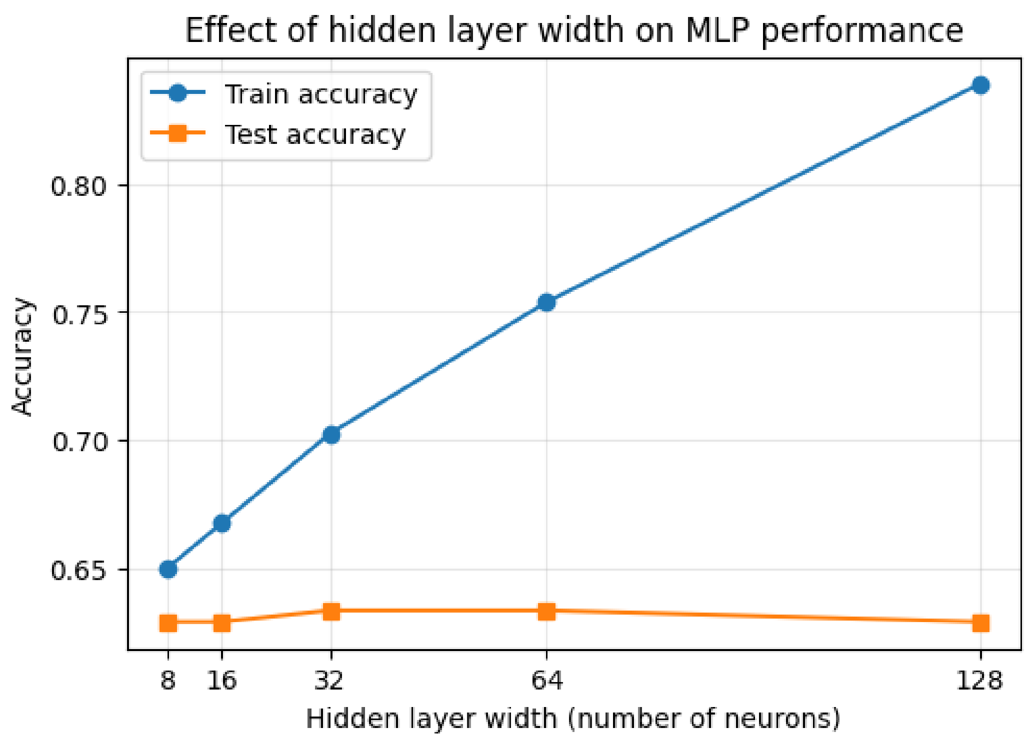
- Training accuracy → how well the model fits data
- Test accuracy → ability to generalise to unseen samples

---

4. Results

4.1 Effect of Hidden Layer Width

📊 *Figure 2 — Train vs Test Accuracy for Different Hidden Layer Widths*



Interpretation

Width	Behaviour
8	Underfitting — model too simple to learn complexity
16–64	Best generalisation — balanced learning
128	High training accuracy but test accuracy plateaus → **overfitting**

This demonstrates the classic **bias–variance trade-off**:

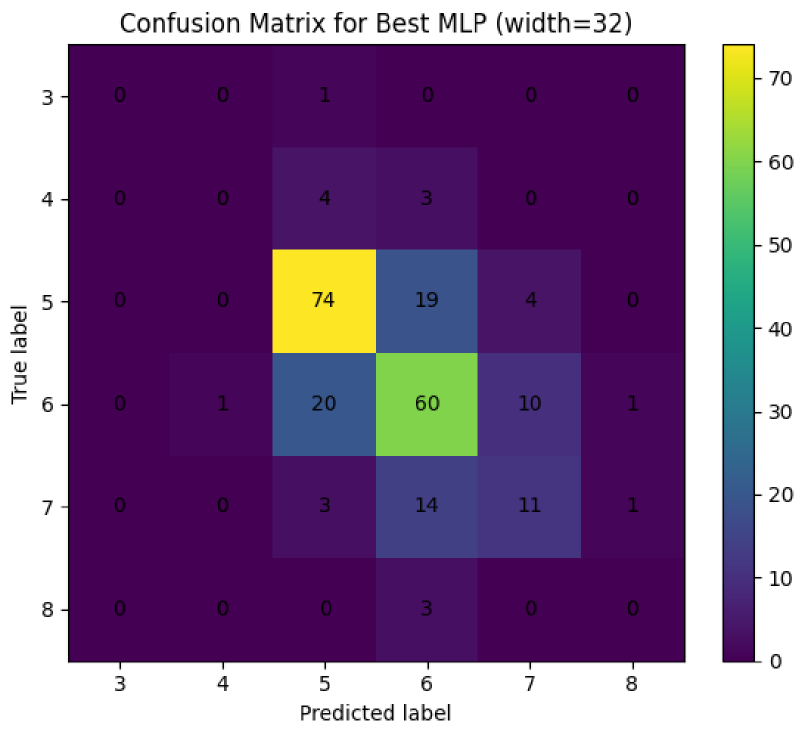
- Too few neurons → high bias, low variance
- Too many neurons → low bias, high variance

A **moderate width** (e.g., 32 or 64) provides the best performance.

4.2 Error Analysis — Confusion Matrix

To better understand model behaviour, we inspect the best-performing MLP:

📊 *Figure 3 — Confusion Matrix for Optimal MLP*



Observations:

- Most confusion occurs between neighbouring quality classes (e.g., 5 ↔ 6)
- Extreme classes (3, 8) harder to classify due to **few training samples**
- Errors reflect **subjectivity in human quality ratings**

---

5. Discussion

5.1 Strengths of the MLP

- Learns complex non-linear relationships
- Outperforms the baseline model
- Improves with increased width **up to a point**

5.2 Limitations

Category	Details
Data	Imbalanced class distribution, limited sample size
Labels	Human-rated quality → noisy and subjective
Model	Only width tested; no depth/regularisation tuning

5.3 Practical Lessons

- **Always try multiple widths** — optimal size depends on task & data
- **Monitor train vs test gap** — prevents overfitting
- **Scaling features** is essential for neural networks

---

6. Ethical & Societal Considerations

Machine-learning predictions may influence winemaking decisions and market value. Therefore:

- Wine quality labels reflect **human taste bias** — must not be treated as objective truth

- Automated grading should **assist**, not replace human experts
- Transparent reporting and open-source code improve **trust and accountability**

---

7. Conclusion

This tutorial showed that:

- MLP accuracy depends strongly on **hidden-layer width**
- Under-sized networks underfit; very large ones overfit
- A **medium-width network** gives the best performance on WineQT

By understanding how architectural choices affect learning, practitioners can design better neural networks for real-world tabular data tasks.

References

- Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). *Modeling wine preferences by data mining from physicochemical properties*. Decision Support Systems, 47(4), 547-553.
- Agrawal, G., & Kang, S. (2018). *Wine Quality Classification with Multilayer Perceptron Neural Network*.
- Scikit-learn documentation: https://scikit-learn.org
- Additional blogs and reading materials consulted for neural-network concepts.

Appendix

- Notebook and dataset available on GitHub : https://github.com/bhanu7959/Machine-Learning-Individual-assignment.git
- Figures auto-generated using the included notebook