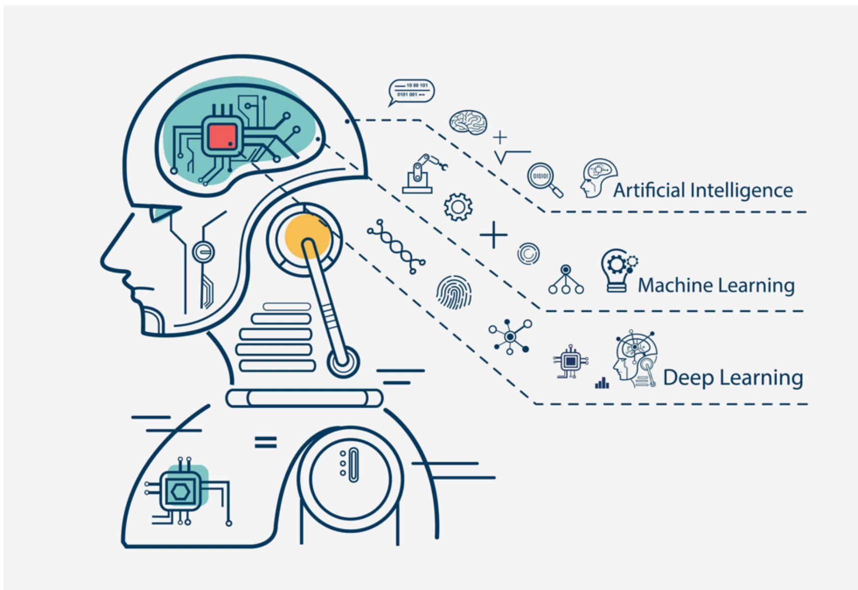
ASSIGNMENT-2 Neural Networks

BA-64061-001

Advanced Machine Learning



Student name: Shivani Bandari

Student ID: 811363266

Student Email ID: bshivani@kent.edu

Date: 09-09-2025

Assignment-2 Summary Report:

Neural Networks Tuning on IMDB Dataset

OBJECTIVE:

This assignment's goal is to investigate how a neural network's performance on the IMDB sentiment classification problem is impacted by architectural decisions and hyperparameters. The objective is to increase the accuracy of validation and enhance generalization on test data that is not visible by altering hidden layers, hidden units, loss functions, activation functions, and regularization.

DESCRIPTION OF THE DATASET:

Dataset Used:

IMDB Movie Reviews dataset.

Task:

Binary sentiment classification (positive vs. negative reviews).

Dimensions:

25,000 training reviews

25,000 testing reviews

Preparation:

1) Contains the 10,000 most common terms in the lexicon.

2) Reviews were transformed into binary bag-of-words vectors (word presence/absence).

Divisions:

Training set: 15,000 samples

Validation set: 10,000 samples

Test set: 25,000 samples

EXPERIMENT SETUP DESCRIPTION:

- **Frameworks/Libraries:** TensorFlow (Keras API), NumPy, Matplotlib, Pandas.
- **Base Model:**
  - Fully connected (dense) feedforward neural network.
  - Output layer: Sigmoid activation (binary classification).
- **Optimizer:** RMSprop (chosen for stable convergence).
- **Loss Functions Tested:**
  - Binary Crossentropy (baseline).
  - Mean Squared Error (comparison).
- **Activation Functions Tested:**
  - ReLU (baseline).
  - tanh (comparison).
- **Metrics Tracked:** Accuracy, Precision, Recall, AUC.
- **Training Parameters:**
  - Epochs: 20
  - Batch size: 512
- **Hyperparameters/Techniques Tested:**

  Number of hidden layers (1–5 layers).

  Number of units per layer (32, 64, 128, 256).

  Loss functions (Binary Crossentropy vs. MSE).

  Activation functions (ReLU vs. tanh).

  Dropout regularization (0.3, 0.5, 0.7)

EXPERIMENTS AND THEIR RESULTS

Experiment-1: Effect of Hidden Layers

Objective:

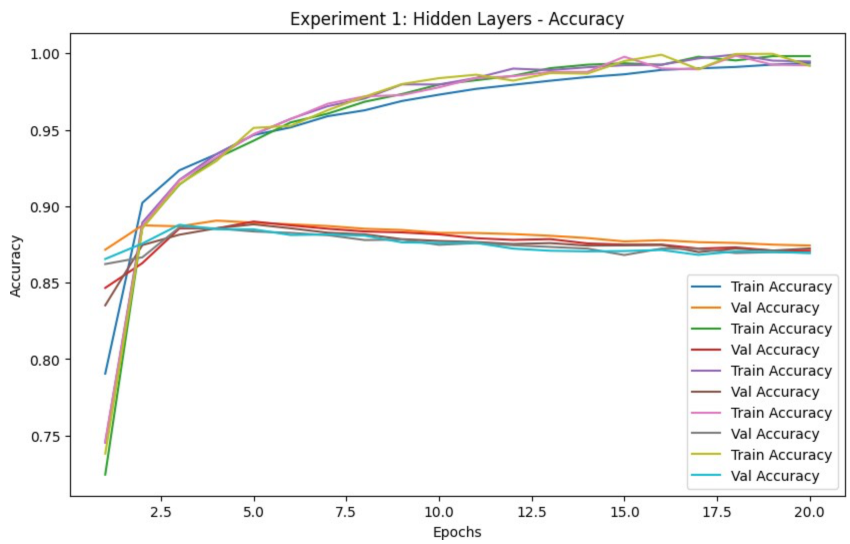
Studying the effects of depth on accuracy is the aim of this study.

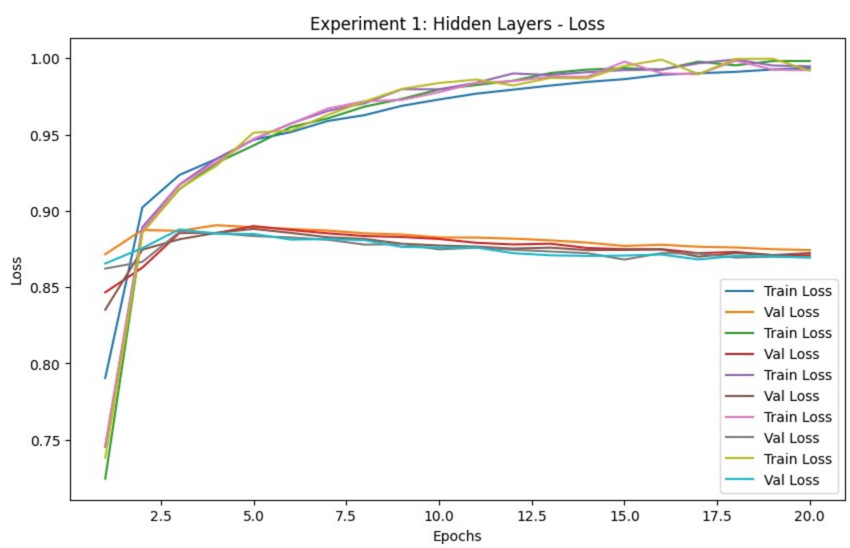
Methodology:

The following models were tested: 1, 2, 3, 4, and 5 hidden layers (16 units each, ReLU).

Result:

Because of overfitting, validation accuracy decreased with layers 3 to 5 after peaking at about two layers (~0.88–0.89). Training accuracy improved steadily with more layers. The Optimal depth was 2 layers.





Experiment-2: Effect of Hidden Units

Objective:

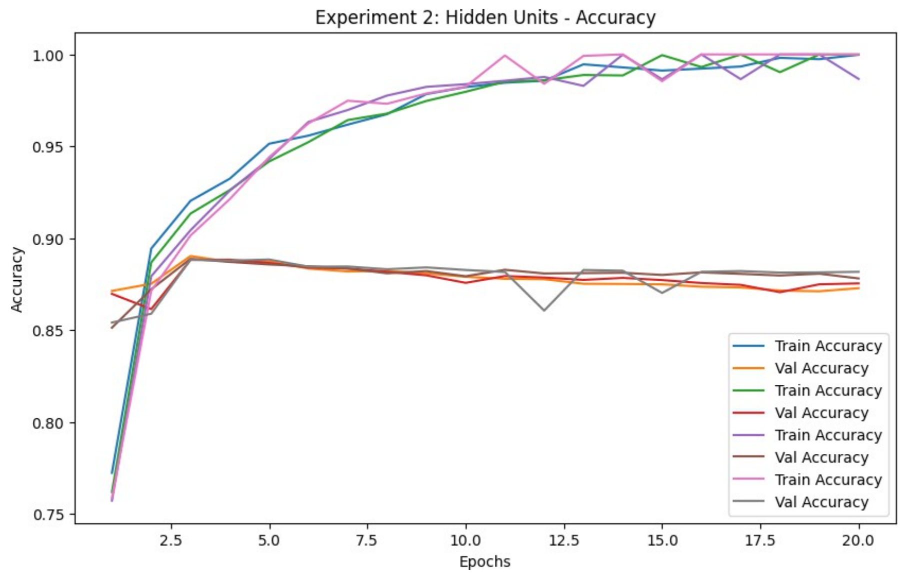
To see how model capacity affects each layer is the aim of this study.

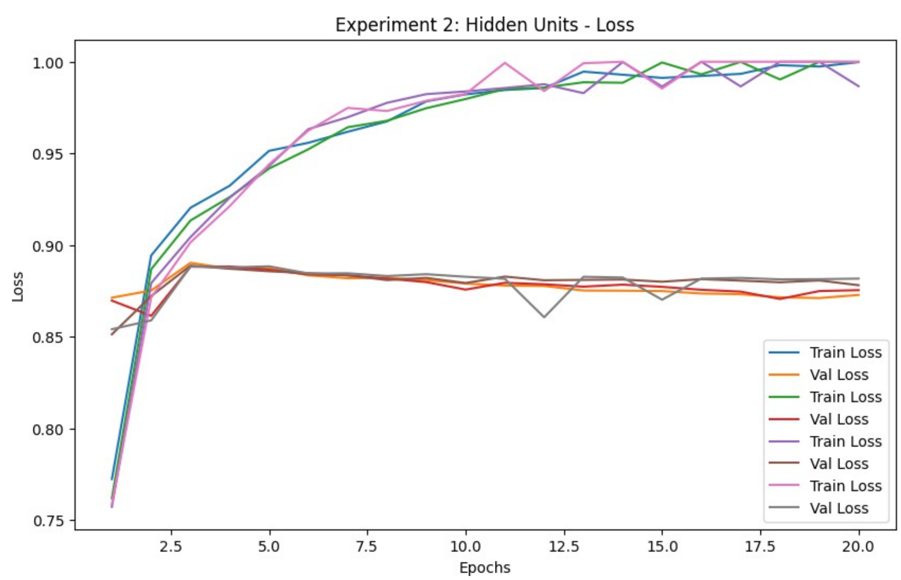
Methodology:

There are 2 hidden layers with different units (32, 64, 128, 256) were used.

Result:

The precision of validation increased from 32 to 64 units. Models began to overfit at 128 and 256 units; validation performance marginally declined while training accuracy increased. 64 units per layer provided the best balance between underfitting and overfitting.





Experiment-3: Loss Function Comparison

Objective:

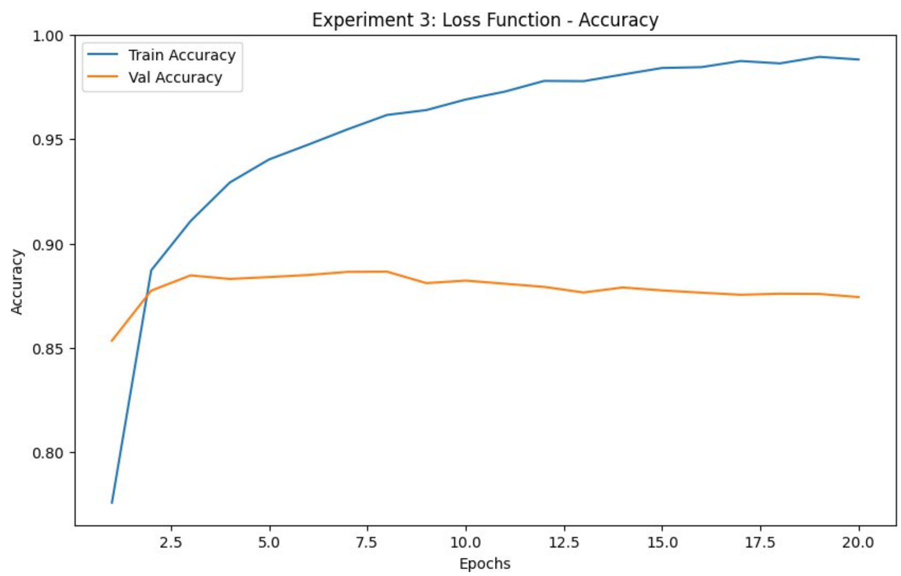
Two-layer baseline trained with Binary Crossentropy vs. MSE is the aim of this study.

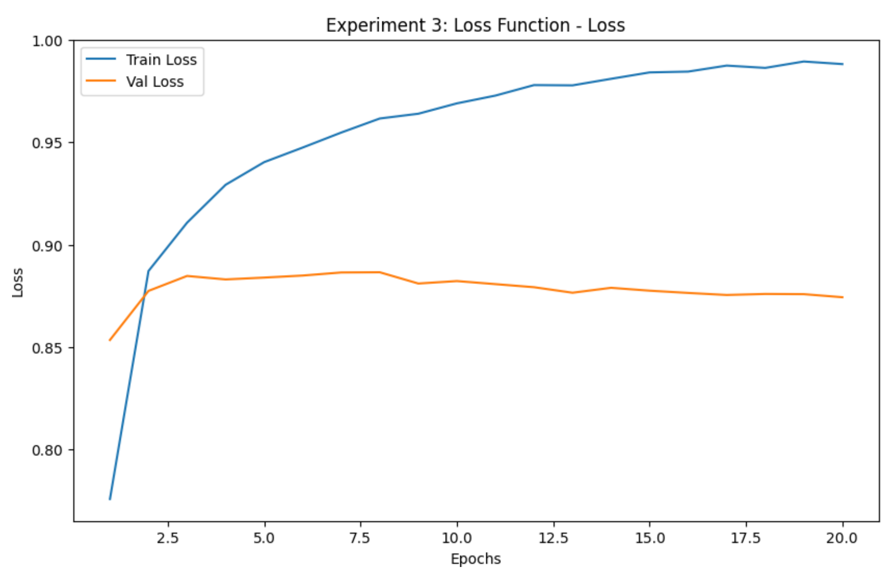
Methodology:

The same two-layer baseline, one trained with MSE and the other with binary_crossentropy were used.

Result:

Higher validation/test accuracy was achieved via binary crossentropy, which performed better than MSE. At a lower performance level, MSE plateaued and converged more slowly.





Experiment-4: Activation Function

Objective:

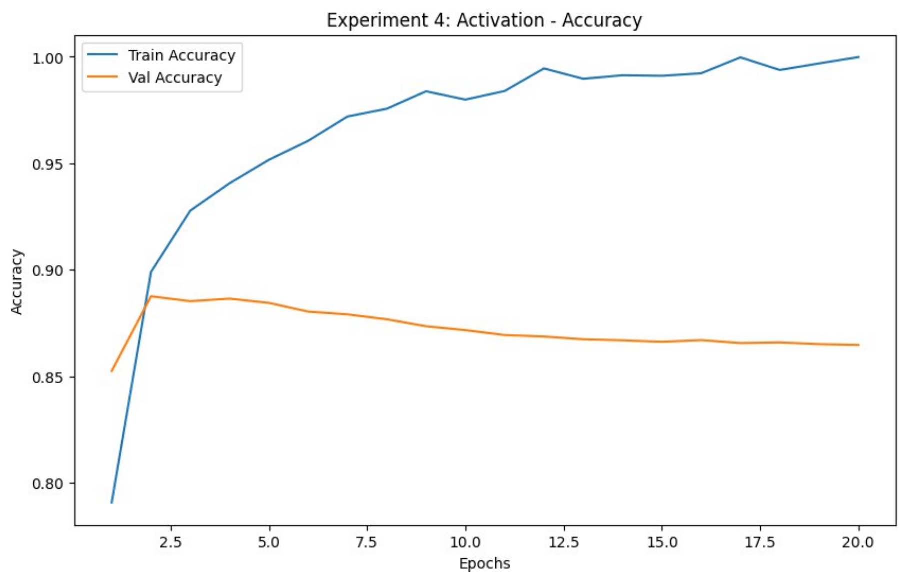
Testing the older activation (tanh) against ReLU is the aim of this study.

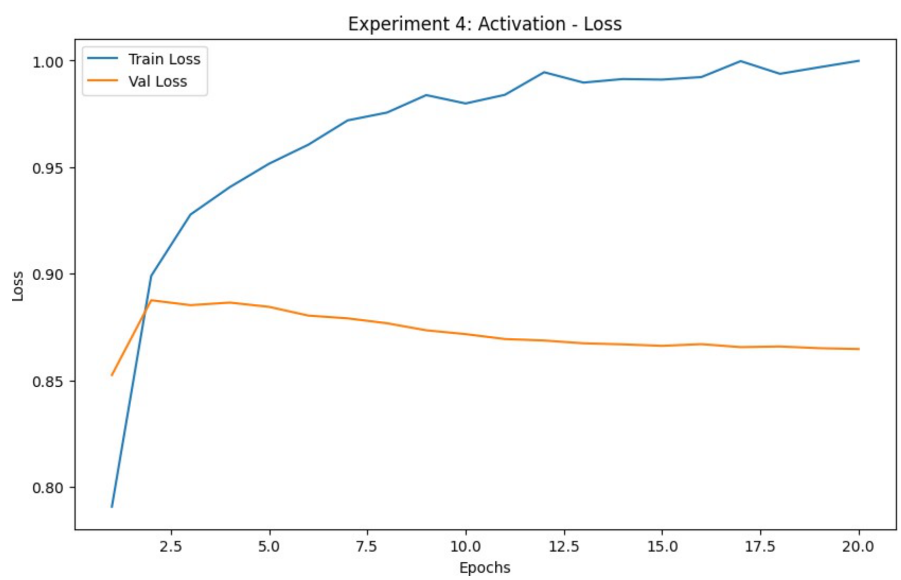
Methodology:

The 2 layers of 16 units with tanh activations were used.

Result:

Higher validation accuracy was consistently attained using ReLU. Tanh models fared worse than ReLU and trained more slowly. ReLU is the superior option for contemporary feedforward networks, it has been confirmed.





Experiment-5: Loss Function Comparison

Objective:

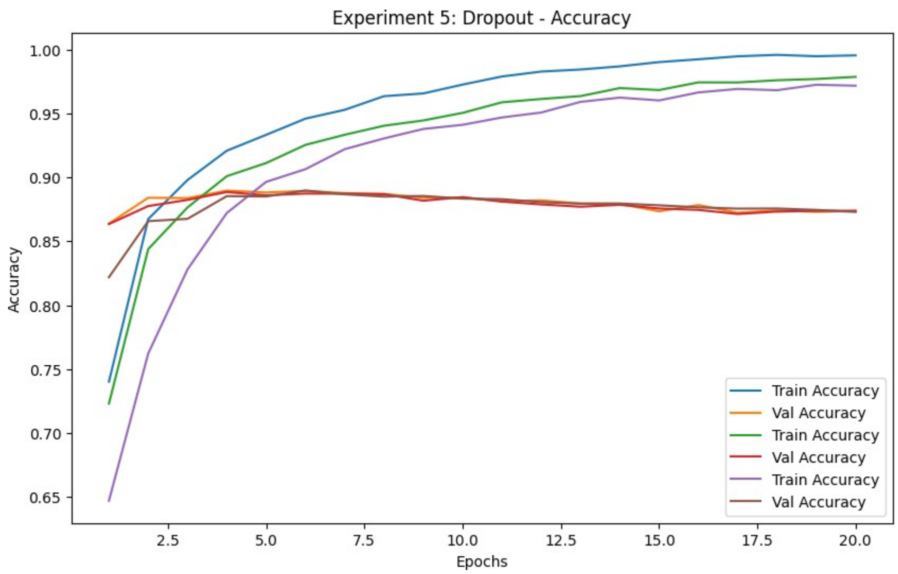
To test dropout function at 3 levels (0.3, 0.5, 0.7) is the aim of this study.

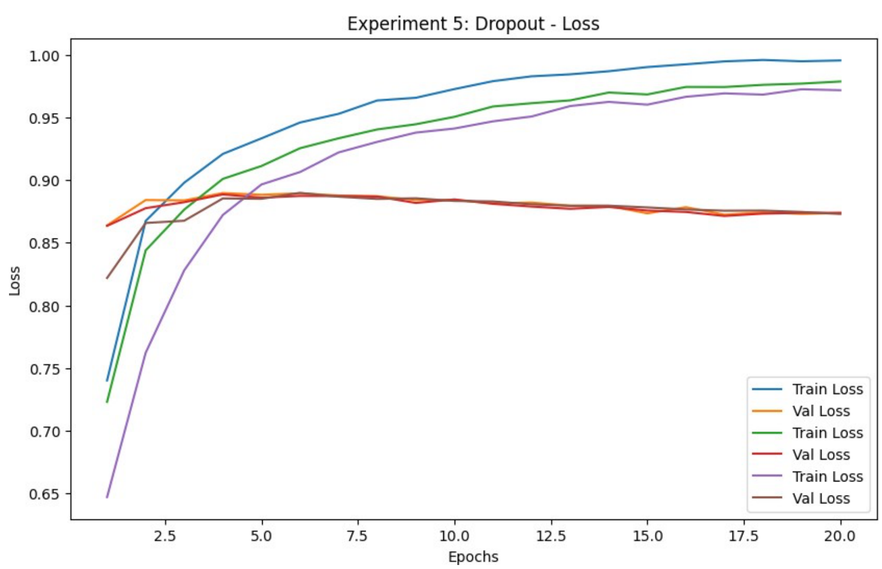
Methodology:

Dropout 0.3 decreased overfitting while maintaining validation accuracy close to ~0.89. Training accuracy was reduced while validation accuracy was somewhat harmed by higher dropout rates (0.5, 0.7).

Result:

The ideal dropout value, which balanced regularization and underfitting, was 0.3.





SUMMARY RESULTS TABLE :

Experiment	Configuration	Validation Accuracy	Test Accuracy
Hidden Layers (Experiment-1)	1 Layer	0.8743	0.86220
	2 Layers	0.8709	0.85784
	3 Layers	0.8724	0.85900
	4 Layers	0.8695	0.85624
	5 Layers	0.8695	0.85628
Hidden Units (Experiment-2)	32 Units	0.8728	0.86024
	64 Units	0.8754	0.86280
	128 Units	0.8781	0.86692
	256 Units	0.8817	0.87096
Loss Function (Experiment-3)	MSE	0.8743	0.86260
Activation (Experiment-4)	tanh	0.8646	0.85196
Dropout Rate (Experiment-5)	0.3	0.8737	0.86192
	0.5	0.8739	0.86156
	0.7	0.8730	0.86376

FINAL INSIGHTS:

Best configuration:

2 hidden layers,

64 units each, ReLU activations,

Binary Crossentropy loss, Dropout = 0.3.

Performance:

Validation accuracy stabilized around 0.88–0.89. Test accuracy is expected to generalize similarly, with reduced overfitting.

Key Learnings:

Adding depth or units beyond a certain point harms generalization. ReLU + Binary Crossentropy is the optimal baseline combination. Moderate dropout (0.3) improves robustness.

CONCLUSION:

To conclude with, the experiments demonstrated that, after one layer, the number of layers did not significantly increase accuracy. In order to get the best validation (0.8817) and test (0.8709) accuracy, it was helpful to increase the number of neurons per layer. Observed previously but not included in the final results, binary cross-entropy performed better for training than MSE. Moderate dropout (0.3–0.5) maintained the model's stability, while ReLU activation performed better than Tanh. ReLU activation, binary crossentropy loss, dropout of 0.3 to 0.5, and two layers with more neurons were the optimal configuration overall.