Tobe Anyigbo
1212073

Homework 3

Question 1

1. Yes. SVC is an implementation of SVM that uses Platt scaling to convert the output of the decision function into confidence probabilities [1]. The probability output can be used for prioritizing threats with the highest threat confidence probabilities.
2.

	Naive Bayes [2]	K-NN [3]	Decision Tree [4]
Training Cost	Low even with large datasets	No training	Expensive for large datasets
Testing Cost	Low. Calculate Posterior Probabilities	Highest	Low. Follow tree rules
Assumptions	Features are independent	Similar instances are closer than dissimilar instances	Data can be split on features
Parameters	Probability	K, distance	Depth
Objective Function	Maximize posterior probability	Majority voting	Minimize class impurity
Optimization	None	None	Splitting class
Expected Dataset Type	Text	Numeric	Text or Numeric

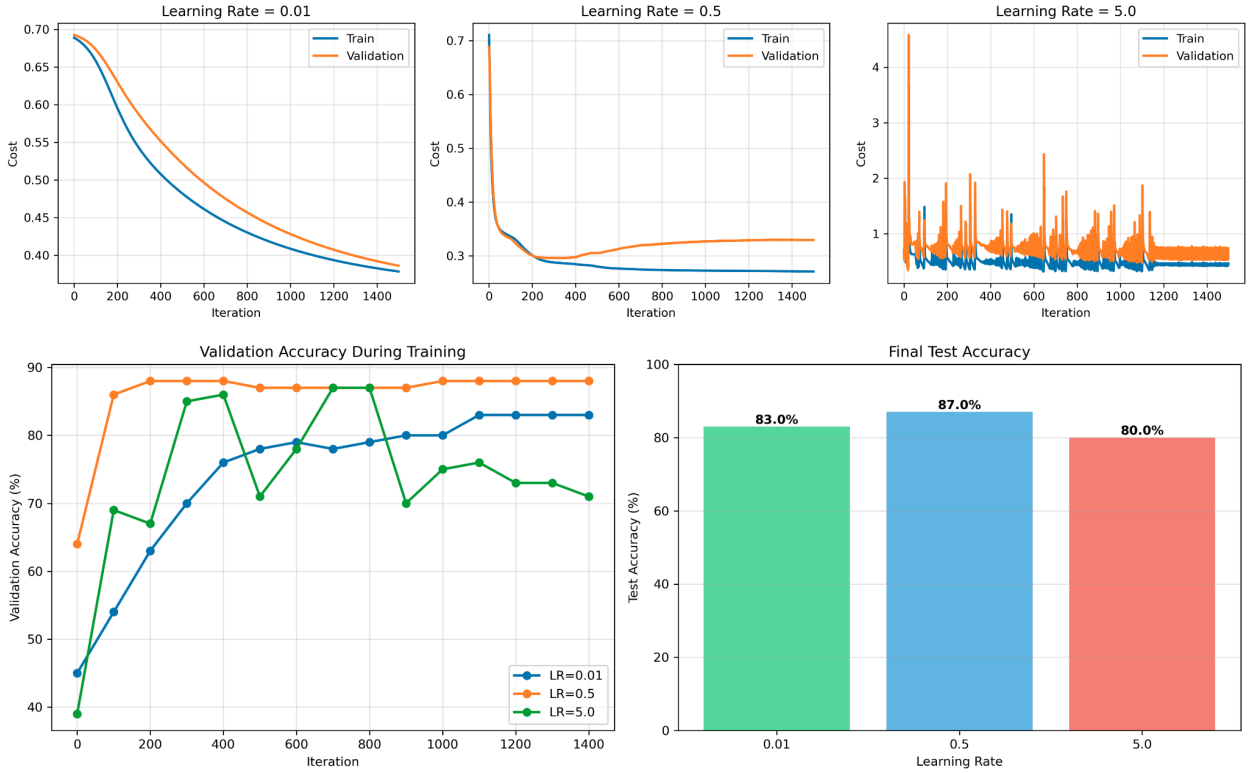
3. CNN and RNN models are both deep learning neural network architectures composed of input, hidden and output layers where the weights of the networks are fine-tuned through backpropagation [5, 6]. CNNs are better suited for image classification and video analysis[5, 7]. RNNs are optimized to handle sequential or time-series data as neurons can retain past information [6, 7]. RNNs like LTSM would be better suited for intrusion detection as they can identify deviations from retained normal patterns.

Question 2

The smallest neural network that can 100% accurately classify the point is a 1 neuron input layer -> 2 neuron hidden layer -> 1 neuron output layer. A 1 neuron hidden layer would not 100% accurately classify the points as the problem is not linearly separable and would result in a max 75% accuracy.

Question 3

1. The 0.5 learning rate performed the best with lowest train and validation cost and highest test accuracy. The low learning rate's slower learning failed to reach optimal weights within the 1500 iterations and the high learning rate has unstable fluctuating losses.
2. With the high learning rate, the weight updates were too large jumping back and forth overshooting the local minimum.
3. The validation set is used to evaluate the model's performance and fine-tune the hyperparameters on unseen data [8].
4. Backpropagation computes the gradient of the loss by applying the chain rule to propagate the error at the output layer backward through each layer to the input layer[9].




```
STEP 4: Generating visualizations...

================================================================================
SUMMARY
================================================================================

Test Accuracy by Learning Rate:
  LR =  0.01 → Test Accuracy = 83.00%
  LR =  0.50 → Test Accuracy = 87.00%
  LR =  5.00 → Test Accuracy = 80.00%
```

Question 4

1. Convolution is a process where a CNN model learns features by scanning the input using a convolution filter to output a feature map [5].
2. CNNs are good at handling image and video data as convolutional operations preserve spatial relationships between pixels which help detect edges and other patterns found in images [5].
3. CNNs transform text into dense vectors of numerical representations of the text [10].
4. Transformers have largely replaced CNNs, RNNs and LSTMs for text data as they solve many of the challenges faced by those models through the combination of parallelized data processing and self-attention mechanisms that more effectively capture long range dependencies [11].

Resources

[1] GeeksforGeeks, “Understanding the predict_proba() Function in Scikitlearn’s SVC,” *GeeksforGeeks*, Aug. 14, 2024. https://www.geeksforgeeks.org/machine-learning/understanding-the-predictproba-function-in-scikit-learns-svc/

[2] GeeksforGeeks, “Naive Bayes Classifiers,” *GeeksforGeeks*, Mar. 03, 2017. https://www.geeksforgeeks.org/machine-learning/naive-bayes-classifiers/

[3] GeeksforGeeks, “KNearest Neighbor(KNN) Algorithm,” *GeeksforGeeks*, Apr. 14, 2017. https://www.geeksforgeeks.org/machine-learning/k-nearest-neighbours/

[4] GeeksforGeeks, “Decision Tree,” *GeeksforGeeks*, Oct. 16, 2017. https://www.geeksforgeeks.org/machine-learning/decision-tree/

[5] GeeksforGeeks, “Convolutional Neural Network (CNN) in Machine Learning,” *GeeksforGeeks*, Dec. 25, 2020. https://www.geeksforgeeks.org/deep-learning/convolutional-neural-network-cnn-in-machine-learning/

[6] GeeksforGeeks, “Introduction to Recurrent Neural Networks,” GeeksforGeeks, Oct. 03, 2018. https://www.geeksforgeeks.org/machine-learning/introduction-to-recurrent-neural-network/

[7] GeeksforGeeks, “Difference between ANN, CNN and RNN,” GeeksforGeeks, Jun. 28, 2020. https://www.geeksforgeeks.org/deep-learning/difference-between-ann-cnn-and-rnn/

[8] GeeksforGeeks, “Computing Gradients with Backpropagation for Arbitrary Loss and Activation Functions,” GeeksforGeeks, Jan. 06, 2025. https://www.geeksforgeeks.org/deep-learning/computing-gradients-with-backpropagation-for-arbitrary-loss-and-activation-functions/

[9] GeeksforGeeks, “Training vs Testing vs Validation Sets,” GeeksforGeeks, Nov. 20, 2021. https://www.geeksforgeeks.org/machine-learning/training-vs-testing-vs-validation-sets/

[10] GeeksforGeeks, “Text classification using CNN,” *GeeksforGeeks*, Jun. 03, 2024. https://www.geeksforgeeks.org/nlp/text-classification-using-cnn/

[11] GeeksforGeeks, “ReLU Activation Function in Deep Learning,” *GeeksforGeeks*, Sep. 30, 2024. https://www.geeksforgeeks.org/deep-learning/relu-activation-function-in-deep-learning/