1. How does the accuracy of classification compare among the single-layer Artificial Neural Network, double-layers Artificial Neural Network, and the traditional supervised machine learning algorithm for the MNIST fashion dataset?

   Accuracy is one of the validation tools to check the model performance on the unseen data.

	Model	Accuracy
0	Model 1	0.8844
1	Model 2	0.8803
2	Model 3	0.8500

   The accuracy changes with the model ideally with the increase of the hidden layers and the number of neurons the models accuracy should increase as the model is able to figure out patterns better but if the process is too complex as compare to the patterns in the data then the model will start to memorize rather than understanding this will result in overfitting of the data and shows drop in accuracy on unseen data

2. What are the differences in terms of training time between the three approaches when classifying the MNIST fashion images?

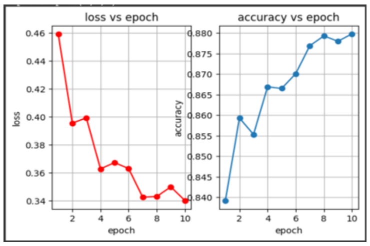
	Model	time
0	Model 1	58.317469
1	Model 2	121.010420
2	Model 3	276.289610

   All the three model they do have different run times for the NN model with 1 hidden layer the run time is the minimum that is because of the lesser number of the parameters to calculate in the model(weights) as compared to the NN model with 2 Hidden layers. But the Most time was taken by the Traditional ML model this is due to the simpler dataset

3. How do the three methods perform in terms of generalization to unseen data? Do any of them show signs of overfitting or underfitting?

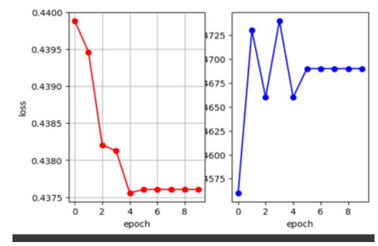
   In Model 1, a neural network with one hidden layer achieved an accuracy of 0.8844 after 10 epochs, with a noticeable stagnation in accuracy improvement after the 8th epoch, indicating that the model had likely reached a plateau in learning. Importantly, there were no significant signs of overfitting, suggesting that it generalizes well to unseen data. In contrast, the traditional machine learning model (logistic regression) exhibited overfitting after just three epochs, leading to a decline in accuracy on unseen data. This rapid onset of overfitting highlights a mismatch between the model's complexity and the dataset's characteristics.

1 hidden layer



2 Hidden layers

Traditional Supervised Model



4. Can you analyze and compare the learning curves of the three approaches? How does the loss and accuracy change over epochs for each method?

   Loss and accuracy are inversely related; as accuracy increases, loss typically decreases. This is evident in the graphs provided. All three models demonstrate that during the first epoch, there is a significant drop in loss and a corresponding rise in accuracy. However, after several iterations, the changes become more erratic, indicating that the model struggles to further learn the data patterns. To improve accuracy, it may be necessary to adjust other hyperparameters, such as the number of neurons or the learning rate.

5. Are there notable differences in the model's performance when dealing with certain classes of fashion items? Do specific approaches excel in classifying certain types of items?

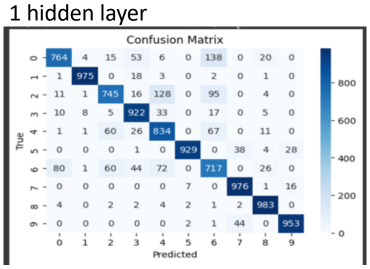
   For model with 1 hidden layer, it works wonders on the simpler images where there were no other classes that were having same attributes but when there were two classes like shirts and t shirts the model got confused and sometime gave the wrong classification.

   For model with 2 hidden layers the model was able to learn a bit more information from the data resulting in higher accuracy with the classes that were doing bad with the 1 hidden layer model but for other classes it started to make wrong predictions this could indicate the overfitting.

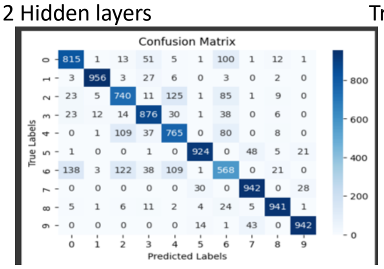
For traditional model it shows the most accuracy, with the same pattern as other 2 models

Overall, all 3 model were performing good with the classes with specific characteristics like sandals and shoes but did not work great with the more complex classes that have similar attributes like shirts and t shirts.

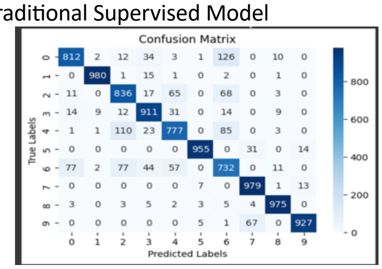
1 hidden layer



2 Hidden layers



Traditional Supervised Model



6. How sensitive are the three methods to hyperparameter tuning? Are there certain parameters that drastically affect their performance on the MNIST fashion dataset?

If we are dealing with the ML model, then the following hyperparameters could be changed.

Learning Rate: With the change in learning rate to a lower learning rate the model will reach a better convergence point that will increase the model's performance, but the time taken to compile the model will increase.

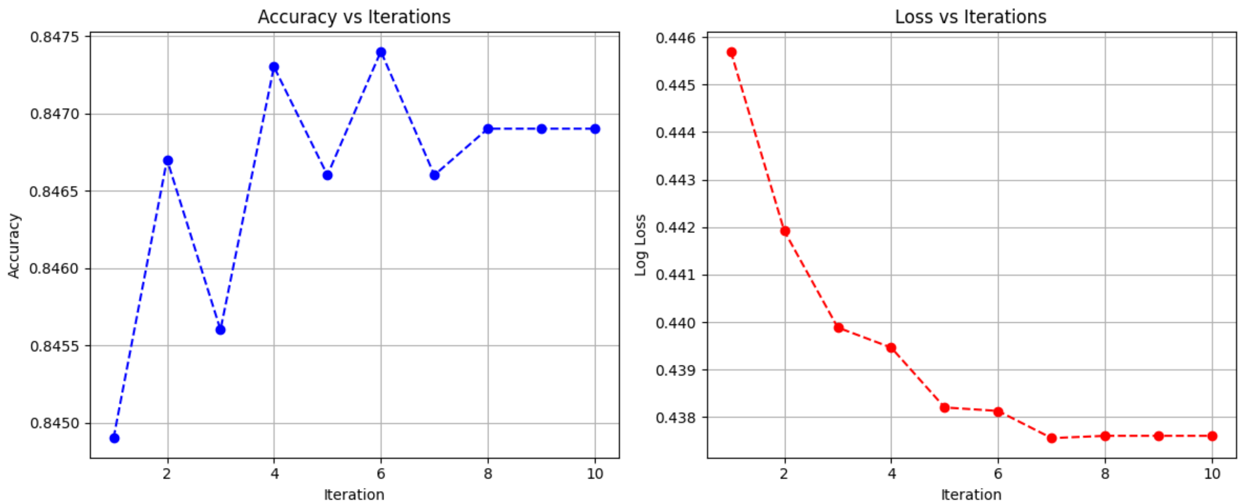
Number of Neurons: the number of neurons have same effect as the learning rate, but the effect will be on the complexity of the model as the more neurons give the model the ability to understand more complex patterns

Activation Function: This makes the model nonlinear and results in a better performance on unseen data. We can use sigmoid, Relu, Leaky Relu, SoftMax depending on the layer and the result required.

Batch Size: Batch size is the number of input that we are processing at a time when training the model this helps reaching the convergence earlier but requires more computation power.

7. What are the limitations of each approach in terms of handling the complexity and nuances of fashion images compared to more advanced techniques?

The limitations of the single layer artificial neural network are that since it has only one layer, it lacks depth, and so it is not able to take as much information from the input and process it. This means that it cannot capture as much details as a NN with more inner layers. The limitation with a double neural network is that it is till limited by the lack of layers compared to deeper neural network. Furthermore, since it is a smaller neural network, it is susceptible to stagnation (lack of growth) or over fitting as seen in the graphs below.



The limitation of the Logistic Regression model is that since it makes the data linear, it risks losing specific non-linear patterns especially in a dataset like fashion mnist which has a lot of details. Compared with more sophisticated models that can handle linear and non-linear information the logistic regression model is on the lesser side.

8. How do the single-layer and double-layered Artificial Neural Networks compare in terms of the depth of features they can extract from the images?

As mentioned before, Single layer neural networks are shallow, so they perform on well on basic patterns, but perform poorly on more complex ones. As for Double layer neural networks, they allow for a deeper understanding because of the larger number of layers that they have. This allows them to take more details from the input data which allows for higher accuracy scores and better handling of unclear images.

9. Can you analyze the decision boundaries created by each approach? Do they handle complex decision boundaries in the dataset differently?

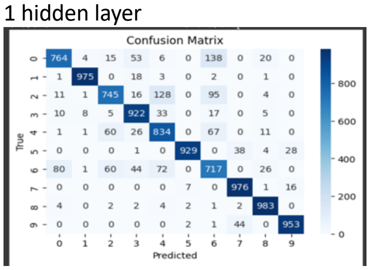
Model with 1 Hidden Layer: This model performed exceptionally well on simpler images where distinct classes had unique attributes. However, it struggled with classes that shared similarities, such as shirts and T-shirts. The decision boundaries for these classes were often blurred, leading to confusion and incorrect classifications, as the model found it challenging to differentiate between the subtle variations in style and design.

Model with 2 Hidden Layers: This model was able to extract more information from the dataset, resulting in improved accuracy for classes that previously underperformed with the 1 hidden layer model. However, it also began making incorrect predictions on some other classes. The decision boundaries for more complex classes remained challenging; for instance, the boundaries between shirts and T-shirts still overlapped significantly, indicating potential overfitting to the training data where the model memorized the training instances rather than generalizing effectively.

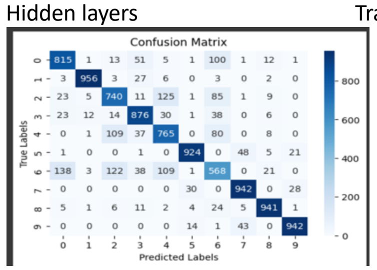
Traditional Model**:** This model achieved the highest overall accuracy, following a similar performance pattern to the neural network models. It excelled in distinguishing classes with specific characteristics but faced challenges with more complex classes that had overlapping attributes. The decision boundaries for items like shirts and T-shirts were still poorly defined, making it difficult for the model to consistently classify these items correctly.

Overall Analysis**:** All three models demonstrated good performance with classes featuring distinct characteristics, like sandals and shoes, where the decision boundaries were clear and well-separated. However, they struggled with more complex classes that shared attributes, particularly shirts and T-shirts. The overlapping decision boundaries for these common classes highlight the need for improved feature representation and potential model refinement to enhance classification accuracy in these cases.

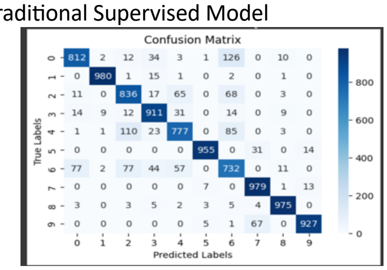
1 hidden layer



2 Hidden layers



Traditional Supervised Model



10. What insights can you draw about the trade-offs between computational complexity, model performance, and ease of implementation when comparing these three approaches?

When addressing a problem, various models can be developed that yield different accuracies, computational complexities, and implementation times. The key is to identify the most suitable approach for the specific scenario. Generally, computational complexity is inversely related to ease of implementation and model performance, up to the point where overfitting occurs.

In many cases, it may be advantageous to opt for a model with slightly lower performance if it significantly reduces implementation time. As computational complexity increases, models often show diminishing returns, with improvements in performance leveling off after a certain number of iterations.

However, in high-stakes scenarios, such as healthcare, where performance is critical, investing in more complex models may be justified. In these situations, the trade-off between increased computational complexity and improved accuracy becomes worthwhile, as even marginal enhancements in performance can have significant implications.