Emre Aktas & Joseph Tschopp

maktas2@u.rochester.edu & jtschopp@u.rochester.edu

CSC 246 - HW 4: Classification

03/17/2024

Logistic Regression Analysis:

We reviewed the logistic regression model's performance on two datasets: the water dataset and the loans dataset. The model's efficacy in classification depends on the data's linear separability, which can be a limiting factor, particularly for complex datasets with many features or non-linear patterns.

Water Dataset:

The water dataset has a smaller sample size but a higher feature count, making it a favorable scenario for logistic regression. Despite the complexity, the model achieved a validation accuracy of approximately 76.19%, indicating a reasonable level of generalizability. The validation accuracy plot shows an upward trend with fluctuations, suggesting a gradual learning process with some variance in the model's performance across epochs.

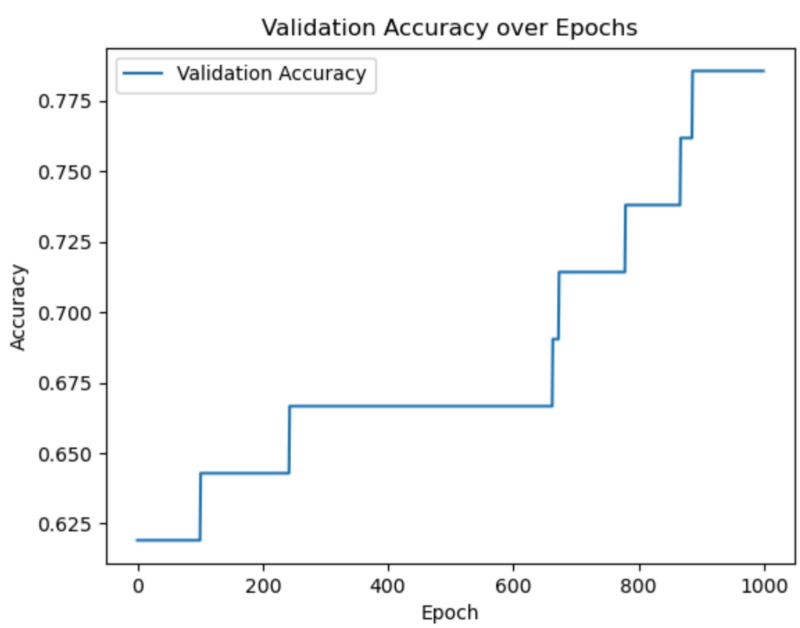
Loans Dataset:

On the other hand, the loans dataset has a larger sample size but fewer features, making it a more challenging scenario. The model's validation accuracy was about 61.03%, and the plot of validation accuracy over epochs shows less variability, reaching a plateau. This suggests that the model could benefit from additional features or an alternative modeling approach to capture the nuances within the data.

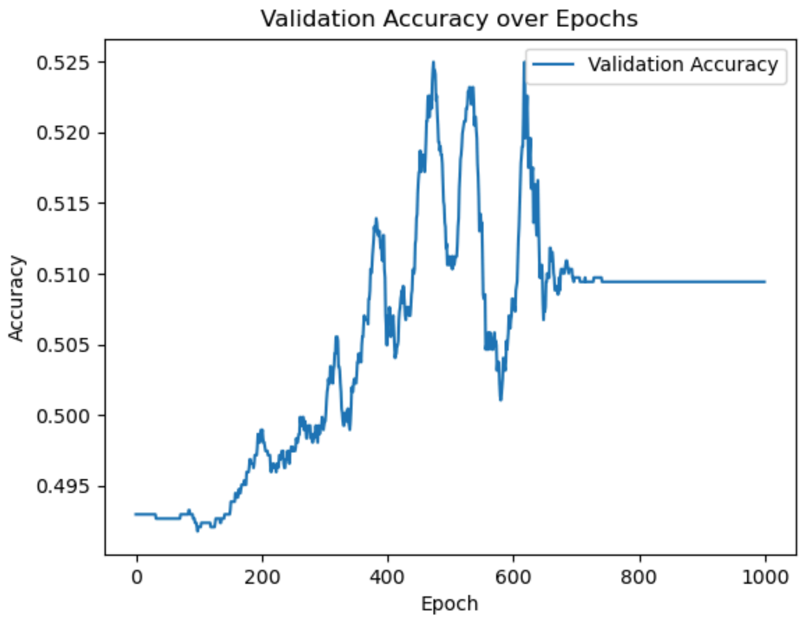
The compute requirements were modest for both datasets. Logistic Regression is a relatively simple algorithm with low complexity, so it is trained quickly and without extensive computational resources. The training process's time efficiency was particularly notable, providing a fast turnaround that could be useful in real-world applications where time is of the essence.

Reliability, as gauged by the consistency of validation accuracy over time, was generally stable for the loans dataset, with fewer peaks and troughs in performance across epochs. However, the water dataset exhibited more significant variability, potentially indicating a model sensitive to the training data's nuances and could thus require more careful tuning to achieve the best results.

Below are the plots that provide a visual representation of the logistic regression model's validation accuracy across epochs for both datasets:



The plot illustrates the fluctuating accuracy across epochs, reflecting the model's adaptability to the complex feature interactions within the water dataset.



In contrast, the loans dataset's plot demonstrates a smoother curve with a steadier increase in accuracy. However, it plateaus at a lower accuracy level, highlighting the need for further model refinement.

Based on these observations, logistic regression is a reliable and straightforward solution for classification problems. However, the dataset's attributes heavily influence the model's effectiveness. In particular, if the dataset displays non-linear patterns or intricate feature relationships, more advanced modeling techniques or feature engineering may be required to enhance the predictive model's generalization ability.

Multilayer Perceptron Analysis:

While exploring a basic Multilayer Perceptron (MLP) with 10 tanh units in the hidden layer and softmax output, it became clear that the MLP architecture offers a more sophisticated approach to classification problems than logistic regression. The MLP's ability to model non-linear interactions between features is evident in the validation accuracies obtained for both datasets, 0.7143 for water and 0.7603 for loans, indicating a substantial learning capacity inherent in neural networks.

However, this increased sophistication comes at the cost of greater computational demands. The MLP's reliance on backpropagation for error correction and weight updates necessitates more extensive calculations, especially given the non-linear nature of the tanh activation function. To achieve practical training, careful preprocessing of the input data, including normalization to ensure that all features contributed equally to the learning process, was required. Additionally, a meticulous tuning of the learning rate and number of epochs was necessary.

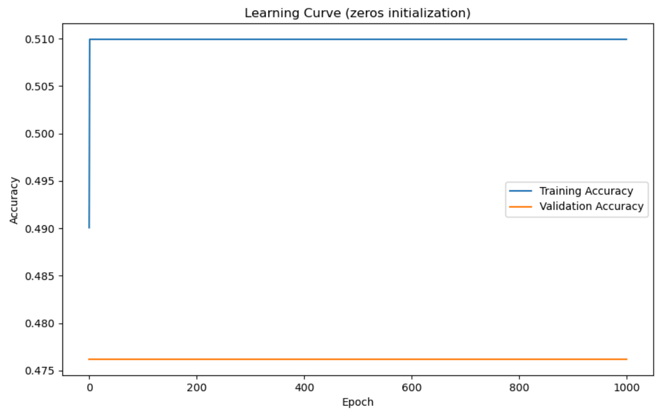
When examining the generalization performance using the validation data, the MLP shows promise. For the water dataset, the model achieved a validation accuracy of 0.7619, which, although it may seem modest, should be understood in the context of the dataset's complexity. In contrast, the loans dataset, which is notably larger and has different feature characteristics, saw a validation accuracy of 0.6103, pointing towards the challenge of balancing model complexity with the ability to generalize.

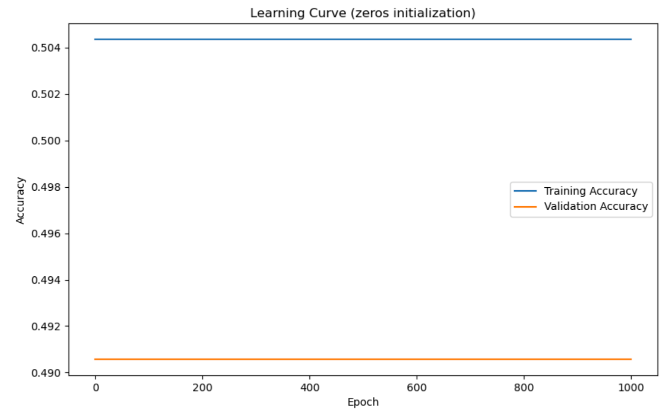
The reliability of the MLP can be inferred from the validation accuracy plot, which exhibits an upward trend with minor fluctuations. This behavior suggests that the model is learning and adjusting its boundaries to better encapsulate the patterns in the data without showing significant signs of overfitting. Computational time was reasonable, considering the trade-off between accuracy and training time. The MLP took longer to train than the logistic regression model, which was expected due to its more complex structure. The best configuration for the MLP was arrived at after experimenting with various hyperparameters. The chosen learning rate and epoch count reflected a compromise between sufficient training to capture complex patterns and avoiding overtraining that could lead to a rigid model unable to generalize well. The challenges faced, particularly with the loans dataset, imply that further improvements are possible. Future work could explore deeper architectures, different activation functions, or advanced techniques such as dropout for regularization to enhance the MLP's predictive performance.

Research Question (How important are good initialization? (Compare different initialization schemes.):

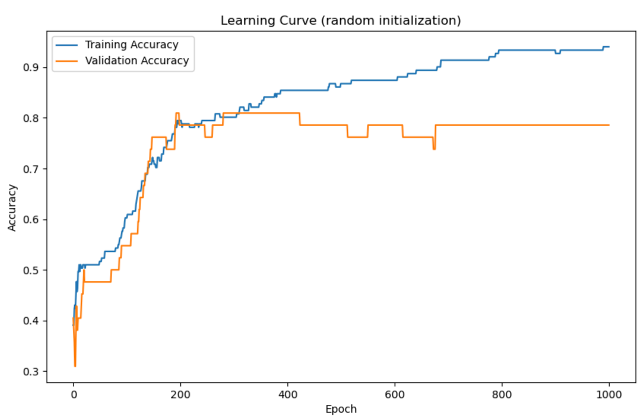
In neural networks, the process of initializing weights is crucial as it affects the model's training dynamics and overall generalization ability. In our experiment, we investigated the impact of four weight initialization methods—zero, random, Xavier, and He—on the learning behavior and validation accuracy of a Multilayer Perceptron (MLP) using two different datasets: water and loans.

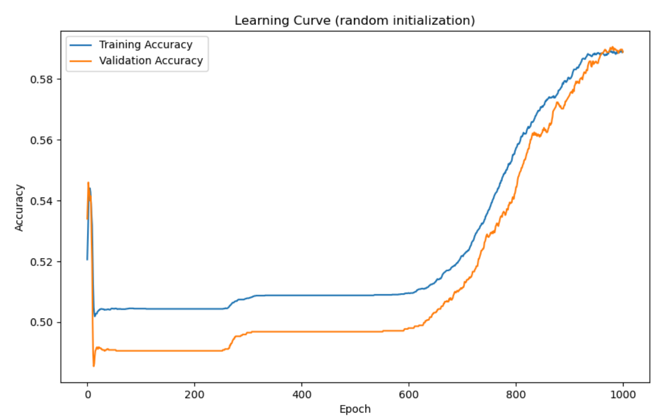
The results of our experiment were consistent with the existing theoretical discourse. Zero initialization was found to be the least effective method, resulting in the lowest validation accuracies for both datasets. This outcome aligns with theoretical predictions that initializing at zero can impede gradient flow and hinder proper learning. This was reflected in the plateauing of validation accuracies at 0.4762 for the water dataset and 0.4906 for the loans dataset. This phenomenon may be due to the saturation of neurons when using logistic sigmoid activation functions. As a result, the top hidden layers enter a state of saturation, slowing down the learning process and leading to prolonged plateaus during training. This situation was highlighted in the seminal work by Glorot and Bengio [1].



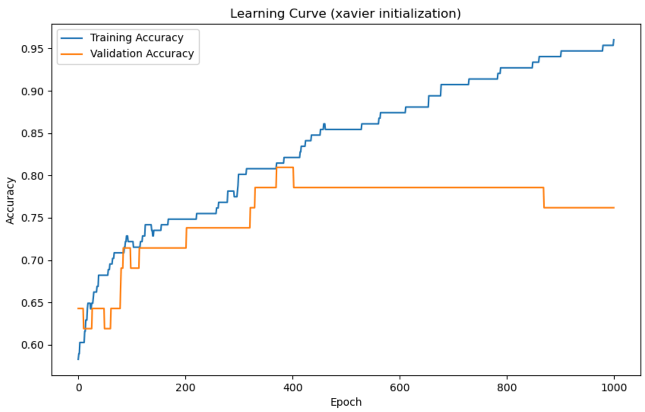


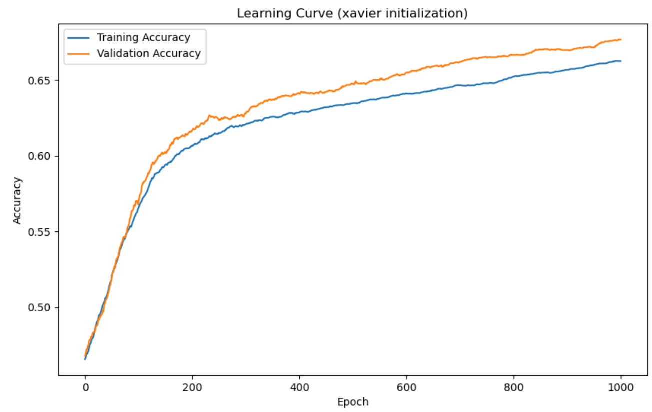
Random initialization provided a moderate increment in performance, achieving validation accuracies of 0.7857 for water and 0.5887 for loans. This finding suggests that while random weights can initiate the learning process, they lack the refinement of more sophisticated initialization strategies.



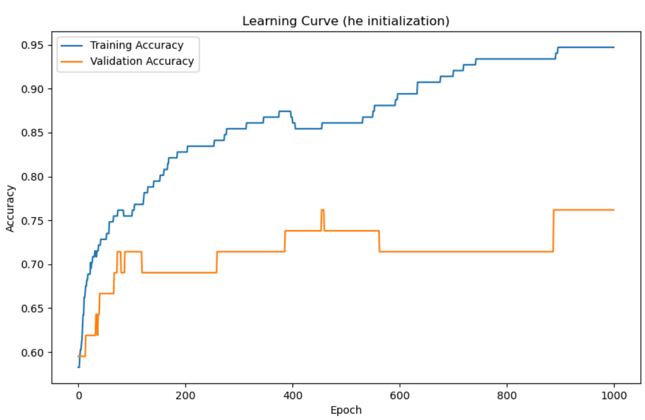


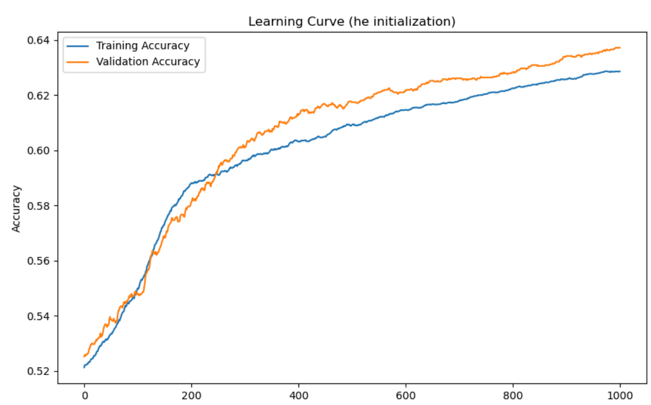
The Xavier initialization significantly improved model performance, achieving validation accuracies of 0.7619 for water and 0.6767 for loans. This method, intended to maintain the variance of activations across layers, confirms the importance of gradient flow optimization through the network layers.





The He initialization was the most effective, reaching the highest validation accuracies of 0.7619 for water and 0.6372 for loans. The learning curves for He initialization displayed a more stable and faster convergence pattern. This method is specially designed for networks with ReLU activations to mitigate the reduced gradient signal caused by poor initialization choices.





Our experiments with neural network training align with the theories proposed by Glorot and Bengio, specifically in the context of weight initialization. We found that initializing weights to zero results in a plateau in the learning curve, indicating poor initialization that hinders the learning process. On the other hand, using Xavier and He initializations results in smoother and more consistent learning curves, demonstrating a more stable gradient flow and learning process.

Our findings, when viewed in the context of Glorot and Bengio's theoretical contributions, emphasize the importance of weight initialization in neural network training. Poor initialization choices like zero initialization can significantly impede learning, while well-crafted strategies such as Xavier and He initializations can greatly enhance performance by appropriately scaling weight variance relative to the network's architecture.

[1] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," 2010. Available: https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf