DS 5220: Supervised Machine Learning and Learning Theory
Summer, 2024

Dr. Fatema Nafa
In Class Work 6

Deadline: 07/ 10/ 2024. 11:59 PM

Learning Objectives: Build a simple neural network model to predict insurance purchase decisions based on customer features. This exercise will guide you through:

1. Data preparation and preprocessing
2. Implementing a basic neural network architecture
3. Training the model using gradient descent
4. Evaluating model performance
5. Visualizing results and model behavior

Students will gain hands-on experience with fundamental concepts of neural networks, including forward propagation, backpropagation, and optimization techniques. The exercise will also cover data handling, model interpretation, and basic hyperparameter tuning.

1. What is the impact of the "affordability" feature on insurance purchase decisions? How might you quantify its importance compared to age?
2. The current model uses only two features. What other features might be relevant for predicting insurance purchases? How would you go about collecting and incorporating this additional data?
3. Experiment with different activation functions (e.g., ReLU, tanh) instead of sigmoid. How do they affect the model's performance and training speed?
4. The current model uses a single neuron. How might increasing the number of neurons or adding hidden layers impact the model's ability to capture more complex patterns?
5. Investigate the effect of different initialization strategies for weights and biases. How do they influence the convergence of the model?
6. The learning rate is set to 0.5 in the custom implementation. Experiment with different learning rates and learning rate schedules. How do they affect the training process and final results?
7. Implement k-fold cross-validation. How does this change your assessment of the model's performance and generalization ability?
8. The current model uses binary cross-entropy loss. Research and implement other loss functions. Are there any that might be more appropriate for this problem?
9. Analyze the misclassified instances. What patterns do you notice, and how might you adjust the model to address these errors?

10. Implement regularization techniques like L1 or L2 regularization. How do they impact the model's performance and prevent overfitting?
11. The current implementation uses a fixed number of epochs or a loss threshold. Implement early stopping based on validation set performance. How does this affect training time and final model quality?
12. Experiment with different optimization algorithms (e.g., SGD, Adam, RMSprop) instead of basic gradient descent. Compare their convergence rates and final performance.
13. How would you handle class imbalance if the dataset had significantly more non-purchasers than purchasers of insurance?
14. Implement a method to interpret the model's decisions, such as SHAP values or LIME. What insights can you gain about the model's decision-making process?