DASC 41103
Machine Learning
Project 1
Trevor Packan and Colin Partridge

1.
    a. Feature scaling is important because it is important for features to contribute equally. Standardizing features leads to more uniform gradients. It also improves model performance as the model is able to be more accurate and stable.
2. Explain the difference between batch gradient descent and stochastic gradient descent.
    a. Batch Gradient Decent
        i. Uses the entire training data set to compute the gradient of the cost function and update weights.
        ii. While the steps can be smooth, the computation becomes slow for large datasets since all samples must be processed for every update.
    b. Stochastic Gradient Decent
        i. Updates weights after each training instead of the whole dataset.
        ii. Faster and more frequently updated, which is better for larger datasets.
        iii. Noisier updates, which can cause the path of optimization to be less reliable.
3. Why does scikit-learn Perceptron and Adline algorithms outperform book code?
    a. Scikit-learn algorithms have already been heavily developed by subject experts, making the code highly optimized. Scikit-learn has advanced configuration options such as learning rate adjustment, regularization and early stopping.
    b. Scikit-learn includes built in tools for scaling and encoding features, giving better results due to a higher consistency of input data.
    c. Scikit-learn is better at cross validation, preventing overfitting.
4. Compare the decision boundaries of logistic regression and SVM
    a. Both models yield nearly identical decision boundaries. This is to be expected if the data is well separated and features are standardized.
    b. The logistic regression boundary is determined by maximizing log-likelihood. While the SVM boundary is determined by maximizing the margin between classes.
    c. Both models achieve the same test accuracy of .8455.
5. What is the role of regularization in preventing overfitting?
    a. Overfitting occurs when the model learns from the training data and the random noise. This causes the model to perform well on the training data but poorly on new data.
    b. Regularization keeps the model weights small. This makes the model focus on the strongest patterns
    c. In the project, we performed a sweep over values of the parameter C. C increased regularization strength, encouraging simpler models and less overfitting.
6. Vary the C values of the scikit-learn LogisticRegression and linear SVC models with [0.01, 1.0, 100.0].
    a. Low C (0.01)
        i. Strong regularization, penalizing large weights.
        ii. Models are constrained - less likely to overfit.
        iii. Lower accuracies
    b. Medium C (1)
        i. Balances bias and variance.
        ii. Improves accuracy, allowing the model to be more flexible without overfitting.
    c. High C (100)
        i. Weak regularization, allowing the model to fit more closely to the training data.
        ii. Accuracy reaches its maximum but with menial gain in context.
    d. Both Logistic Regression and LinearSVC perform well across C values but excessive regularization hurts performance due to over simplification. Reducing regularization boosts accuracy up to a point to a point in which the overfitting does more harm.