**Why is feature scaling important for gradient-based algorithms?**

* When features are on different scales, algorithms can struggle to converge in a timely manner due to features with larger ranges dominating the gradient. Feature scaling allows variables to not be normalized and more balanced. This improves accuracy and interpretability, in addition to a faster convergence time.

**Explain the difference between batch gradient descent and stochastic gradient**  
**descent.**

* **Batch:**
* Uses the entire dataset to compute the gradient before updating parameters.
* Pros: Stable convergence, smooth loss curve.
* Cons: Quickly becomes computationally expensive with more data.
* Stochastic:
* Uses one random sample at a time to compute the gradient and update parameters.
* Pros: Much faster and allows training on larger datasets.
* Cons: Samples can be noisy, and the model requires more careful tuning of the learning rate.

**Why does scikit-learn Perceptron and Adline algorithms outperform book code. Research and develop an informed answer.**

* skitlearn adds regularization and optimization to make code more efficient (than our custom adaline) - from class, need to add more

**Compare the decision boundaries of logistic regression and SVM.**

**What is the role of regularization in preventing overfitting?**

* Overfitting occurs when a model fits training data too closely and fails to generalize on new data. Regularization combats this by penalizing the number of features. By doing this, you limit how complex you can make the model, resulting in a model that is not overly tuned to the training data.

**Vary the C values of the scikit-learn Logistic Regression and linear SVC models**  
**with [0.01, 1.0, 100.0]. Discuss the impact.**

* Logistic Regression:
  + Results:
    - Accuracy was 0.83 for all three C values
    - Precision, recall, and F1 scores were very consistent
    - Confusion matrices very consistent
  + Why:
    - We used StandardScaler during preprocessing, which ensures all features are on the same scale. This makes regularization more balanced across features, so changing the C value doesn’t drastically shift the model’s behavior.
    - Our model also learned a stable decision boundary — meaning it consistently separates the classes well without needing extreme coefficient values or heavy regularization.
    - Finally, the class imbalance in our dataset had a stronger influence on performance than regularization did. Since the minority class is underrepresented, tuning C alone wasn’t enough to improve its recall or precision.
* SVM:
  + Results:
    - Accuracy was around .85 for all three C values with both the rbf and linear kernels
  + Why: