Title: PS2

Author: Vitor Hugo Silva dos Santos

Instructor: Dr. Tan

Class: Applied Machine Learning in Economics (ECON-6930-01)

Date: October 14, 2024

1   Introduction

In this assignment, I worked on implementing linear classifiers using hinge and softmax loss functions. The goal was to test how well these classifiers work on a spiral dataset. I also used regularization to prevent overfitting and applied gradient descent to improve the model over time.

2   Methodology

2.1   Score Function

The score function is a simple calculation. It multiplies the input features by the weights and adds a bias. This can be written like this:

$$\text{scores} = XW + b$$

Where X is the data, W is the weights, and b is the bias. This is the basic step used before calculating losses.

2.2   Hinge Loss

Hinge loss is mostly used in SVMs. It calculates how far off the prediction is from the correct label, but only when it's really off. The formula is:

$$L_{\text{hinge}} = \max(0, 1 - y \cdot \text{score})$$

I also used L2 regularization to keep the weights from getting too big:

$$\text{L2 Regularization} = \lambda \sum W^2$$

2.3 Softmax Loss

Softmax loss is used when there are multiple classes. It calculates the probability of each class and measures how far off it is from the true label:

$$L_{\text{softmax}} = -\log\left(\frac{\exp(\text{score}_{y_i})}{\sum_j \exp(\text{score}_j)}\right)$$

I used L2 regularization here too, to keep the model from overfitting.

2.4 Gradient Descent

I used gradient descent to make the model better by adjusting the weights and bias. In each step, it calculates how much the loss changes for small changes in W and b, and then updates them:

$$W = W - \eta \cdot \nabla_W L$$

$$b = b - \eta \cdot \nabla_b L$$

Where η is the learning rate, which controls how big each update is.

3 Results

I tested the model on a spiral dataset. Here are the results from the loss functions, both with and without regularization.

3.1 Task 1: Score Calculation

The calculated scores were:

$$\text{Scores} = \begin{bmatrix} 0.66 & 0.94 \\ 0.6 & 0.7 \end{bmatrix}$$

3.2   Task 2: Loss without Regularization

- **Hinge Loss**: 0.8067

- **Softmax Loss**: 1.0331

3.3   Task 3: Loss with L2 Regularization

- **Hinge Loss (with L2 regularization)**: 0.8327

- **Softmax Loss (with L2 regularization)**: 1.1291

3.4   Task 4: Gradient Descent Optimization

For the hinge loss function:

$$\text{Optimized Hinge Weights} = [0.506, 0.601], \text{ Optimized Bias} = 0.42$$

For the softmax loss function:

$$\text{Optimized Softmax Weights} = \begin{bmatrix} 0.244 & 0.601 \\ 0.088 & 0.735 \\ 0.651 & 0.137 \end{bmatrix}, \text{ Optimized Bias} = [0.129, 0.205, 0.266]$$

4   Conclusion

I learned a lot about hinge and softmax loss functions and how gradient descent helps improve the model. Regularization made a big difference in keeping the model from overfitting. The hardest part was finding the right balance with the learning rate and regularization. This assignment helped me understand these concepts better, and it was great to apply them to a real dataset.