ECON 6930 — Problem Set 2

Nicola Pipoli

October 14, 2024

1 Introduction

The goal of the assignment was to implement a linear classifier using both SVM and Softmax classifiers. Furthermore, L2 regularization was implemented to prevent overfitting, and gradient descent to minimize the loss functions, ensuring optimal model training.

2 Methodology

- Score Function: The score function was computed using matrix multiplication of the input data X and weights W, followed by the addition of bias b. This provided the raw class scores before applying the loss functions.
- Hinge Loss and Softmax Loss: Hinge loss was used for the SVM classifier, calculating the margin-based loss. Softmax loss involved applying the softmax function to convert raw scores into probabilities and calculating the cross-entropy with the true labels.
- Regularization: L2 regularization was added to both loss functions to reduce weights given to large values, reducing overfitting and improving generalization.
- Gradient Descent: gradient descent was implemented to minimize the loss function. Gradient descent looks at the vector of partial derivatives and aims to iteratively train the model by with the goal to minimize the loss function.

3 Results

Below are represented the output of the codes generated, starting with a picture of the dataset showing its non-linearity, as well as results for the score shape (300, 3) and SVM Loss of 1.997042089440211 and Softmax Loss of 1.097643146638078

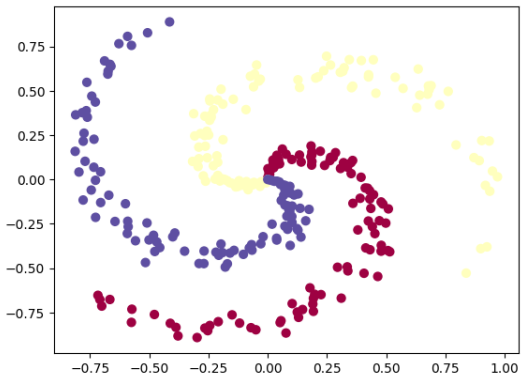

Figure 1: Data Representation Showing Non-Linearity.

.

4 Conclusion

In conclusion, the primary goal of this assignment was to implement and understand key machine learning concepts such as the score function, hinge loss (SVM), softmax loss, L2 regularization, and gradient descent. Through the implementation of both the SVM and softmax classifiers. Regularization helped to reduce overfitting by penalizing large weights, resulting in a more generalized model.Gradient descent is then used to train the model and improve its accuracy. However it is duly noted that gradient descent optimization can at times be counterproductive, especially when the learning rate is set improperly, hence having the model stuck in a local minimum.

The biggest challenge was ensuring proper matrix dimensions and aligning shapes for operations like matrix multiplication, as the code returned errors related to matrix dimensionality. Another complex code implementation related to inlcuding a code that would "print" the values for both the SVM and Softmax loss. Once the code was refined, the model was complete.