Machine Learning Homework 02 – Report

Adam Catto

1 Introduction

The main goal of this assignment is to implement a linear classifier to perform multiclass classification. In order to initialize the relevant weights ω for the classifier, there are two approaches to take:

1. Initialize ω with the first data point of the dataset
2. Initialize ω using linear regression

For approach (2), we are meant to also implement our own linear regression algorithm. Finally, we are to train and validate our implementation using 5 samples from the scikit-learn datasets "Breast cancer" and "Iris".

2 Solution

2.1 Linear Regression Implementation

To do linear regression, we are to solve the least squares error problem for a dataset. We have a system of equations represented as

$$X\omega = \mathbf{y}$$

We can approximate a solution to this by setting an error term $r = X\omega - \mathbf{y}$, and minimize the size of r – this is the least-squares approximation of a solution to the system of equations.

To solve for ω, we take the following derivation steps:

$$X\omega = \mathbf{y} \tag{1}$$

$$\implies X^T X\omega = X^T\mathbf{y} \tag{2}$$

$$\implies \omega = (X^T X)^{-1} X^T\mathbf{y} \tag{3}$$

The