Neural Networks with Iris’

Background

This project utilizes the iris data set. A commonly known data set collected by Edgar Anderson who aimed to “quantify the morphologic variation of Iris flowers of three related species.” There are three species in this set known as setosa, virginica, and versicolor with 50 samples from each, and their associated length and width for sepal and petal. Fisher used this data for the purposes of creating a “linear discriminant model to distinguish the species from each other.” But many others have gone on to use this set like Ronald Fisher, a statistician and biologist that kept developing the linear discriminant model to understand how features can be related or not and help determine classification.

MLP models

The MLP model I created has a hidden layer size of four and only iterates 1000 times using the solver ‘lbgfs’. This just means I have four neurons looking at the data 1000 times using a quasi-Newton solver which lets the solver use approximates for a faster converge. The MLP model I created stacks up pretty well to a logistic regression model. Over multiple train/test splits the logistic regression model can vary anywhere from a ~50% accuracy score to a full 100%, while the MLP has a smaller accuracy score window between ~65% and 100%.

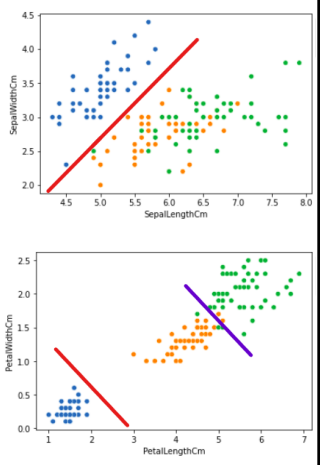
Confusion Matrices and Accuracy

Looking at the two models’ confusion matrices and accuracy scores showed a lot of similarities in deficiencies. While the window of accuracy was much larger for the logistic regression model, both that model and the mlp struggled to accurately place versicolor and virginica flowers. Most errors showed up in the confusion matrix as versicolor being mistaken as virginica, slightly more than vice versa and only a few mistakes setosa as versicolor. I’ve seen other small errors like setosa as virginica, but they don’t happen often. Though once in a while the logistic regression model will completely mix up the classification and have most if not all setosa classified as versicolor. This probably has more to do with the sepal information than the petal seeing as petal information set the setosa in it’s own island. Most issues come from the versicolor flowers being categorized than anything else in both models, and I think in turn, also causes the setosa and virginica to be mixed because of it. If the data was only setosa and virginica I don’t think there’d be as many issues.

ReLU Activation

When calculating the outputs using the weights and biases the output would come out as <0, 0, or >0 which helped determine where the point would be on the plane. By using the ReLU function, it took those numbers that were <0 and made them *exactly* zero. It helps speed up the training period since there aren't any other equations needed when dealing with points on the opposite side, but also can lead to neurons that aren't super useful if their learning is all negative or zero. It can make decision boundaries super useful if the learning goes well and each neuron can focus on one side of the plane to help classification, on the other hand it can make some decision boundaries difficult to use since outputs that are heavily on the opposite side can skew the classification incorrectly.

The Planes I See in My Head But Can't Explain Well, Sorry About That

Setosa vs. The Rest (Red Line)	Setosa is the easiest to separate compared to versicolor and virginica. Their sepal W/L and petal W/L are fairly distinct and don't intermingle as much as versicolor and virginica	Sepal W/L increase linearly with small exceptions and petal W/L aren't that far off since they're secluded with much smaller measurements.
Versicolor/Virginica vs. The Rest (Purple Line)	Unfortunately Sepal W/L for versicolor is much harder to separate considering their similarities to virginica.	The easiest way to determine versicolor and virginica would be to look at their petal W/L as there is only small overlap, but generally versicolor petal W/L is a bit smaller.
	It would be interesting to make a 3D visualization of this data using the petal plot but also adding a third axis for either sepal width or length. If they change the difference at all. The petal information for all three flowers is by far the best way off initial look for determination, but obviously a MLP can't do that initially which is why there's training.	I wonder how the classification would change if it were just petal or just sepal information. It seems like the linear regression model would have an easier time with petal but fail on sepal, while mlp would probably be able to handle both within reason.