Text, letter

Description automatically generated

There is three classes in the data, three species of this Iris flower apparently. There are four features, sepal length, sepal width, pedal length and pedal width. Which appear to be measurements of the flower. The sepal is the outer green petals that protect the flower before it blooms, they form the bud I suppose you would say and therefore you might expect the bigger sepals to have bigger flowers although the data at a cursory glance seems a bit mixed on that assumption. The length seems to be related while the width does not.

There does seem to be some differences between the classes between measurements, for example below you can see the difference in petal width between Setosa and Versicolor is significant with Setosa’s being around .2-.4 and Versicolor’s being around 1.3-1.5 and it is like this for several of the measurements and classes and this would lead me to believe that we would be able to classify these well with the methods we have at our disposal.

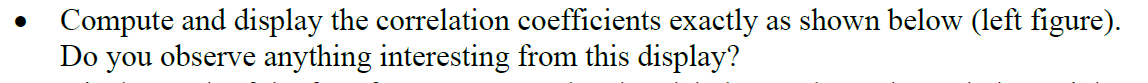
Table

Description automatically generated

Table

Description automatically generated

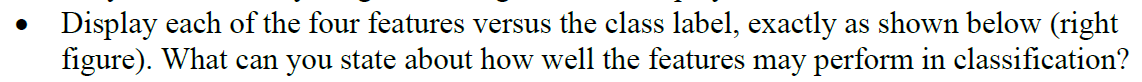
The Between-Class Variance of Versicolor being so low seems alarming. This seems to say that Versicolor is not very different than the other two and therefore difficult to classify. I will bring up IID now because this seems to be the first place where it comes into question. Some of the Within-Class Variances being bigger than that Between-Class Variance makes it seem like you will have a harder time telling different Versicolors apart than telling it from other classes.



I observe something very interesting from this display which is that petal length and petal width are highly correlated, that is a serious problem for IID which is what our classifiers rely on to be effective. The measurements are not independent. Even petal length and sepal length are highly correlated as I suggested earlier. Although petal width and sepal width are not which also supports my earlier observations.

Chart

Description automatically generated



I mean yeah, you can already see how only a couple of them are linearly separable which is going to be a problem for Perceptron and expect some misclassification from Least Squares. Petal Length and Petal Width appear linearly separable between Setosa and Versicolor although not with Virginica and Versicolor.

Graphical user interface, chart

Description automatically generated

Text

Description automatically generated



19 epochs for Batch Perceptron to converge. 0 misclassifications. Too many dimensions to plot.

Text

Description automatically generated



255 epochs. 0 misclassifications.

Text

Description automatically generatedChart, scatter chart

Description automatically generated



Perceptron never converged. 36 misclassifications with Least Squares. Too many dimensions to plot.

Text

Description automatically generated



Perceptron never converges because they are not linearly separable. Least Squares gets 37 misclassifications although that can be modified a bit if you bias it a bit, there’s twice as many samples in the one class than the other which is part of what’s biasing it so far the other way.

Text

Description automatically generatedChart, scatter chart

Description automatically generated



Least Squares has a pretty hard time with this because Petal Length and Petal Width are highly correlated especially between Versicolor and Virginica. 50 misclassifications

Text

Description automatically generatedChart, scatter chart

Description automatically generated