Python Project 2018

Started Thursday 29th March. Initial research on Iris Flower Data Set on Wikipedia <https://en.wikipedia.org/wiki/Iris_flower_data_set>.

The ***Iris* flower data set** or **Fisher's *Iris* data set** is a [multivariate](https://en.wikipedia.org/wiki/Multivariate_statistics) [data set](https://en.wikipedia.org/wiki/Data_set) introduced by the British [statistician](https://en.wikipedia.org/wiki/Statistician) and [biologist](https://en.wikipedia.org/wiki/Biologist) [Ronald Fisher](https://en.wikipedia.org/wiki/Ronald_Fisher) in his 1936 paper *The use of multiple measurements in taxonomic problems* as an example of [linear discriminant analysis](https://en.wikipedia.org/wiki/Linear_discriminant_analysis).

The data set consists of 50 samples from each of three species of *Iris* ([*Iris setosa*](https://en.wikipedia.org/wiki/Iris_setosa), [*Iris virginica*](https://en.wikipedia.org/wiki/Iris_virginica) and [*Iris versicolor*](https://en.wikipedia.org/wiki/Iris_versicolor)). Four [features](https://en.wikipedia.org/wiki/Features_(pattern_recognition)) were measured from each sample: the length and the width of the [sepals](https://en.wikipedia.org/wiki/Sepal) and [petals](https://en.wikipedia.org/wiki/Petal), in centimetres. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

Based on Fisher's linear discriminant model, this data set became a typical test case for many [statistical classification](https://en.wikipedia.org/wiki/Statistical_classification) techniques in [machine learning](https://en.wikipedia.org/wiki/Machine_learning) such as [support vector machines](https://en.wikipedia.org/wiki/Support_vector_machines)[[5]](https://en.wikipedia.org/wiki/Iris_flower_data_set#cite_note-5).