Iris Report

Introduction

The **Iris flower** data set, or Fisher's Iris data set, is a multivariate data set introduced by the British statistician, eugenicist, and biologist Ronald Fisher in his 1936 paper 'The use of multiple measurements in taxonomic problems as an example of linear discriminant analysis'. It is sometimes called Anderson's Iris data set because Edgar Anderson collected the data to quantify the morphologic variation of Iris flowers of three related species. Two of the three species were collected in the Gaspé Peninsula "all from the same pasture, and picked on the same day and measured at the same time by the same person with the same apparatus".

The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

Iris Virginica

Analysis

Sample Data

sepal_length	sepal_width	petal_length	petal_width	species
5.8	4.0	1.2	0.2	setosa
5.1	2.5	3.0	1.1	versicolor
6.6	3.0	4.4	1.4	versicolor
5.4	3.9	1.3	0.4	setosa
7.9	3.8	6.4	2.0	virginica
6.3	3.3	4.7	1.6	versicolor
6.9	3.1	5.1	2.3	virginica
5.1	3.8	1.9	0.4	setosa
4.7	3.2	1.6	0.2	setosa
6.9	3.2	5.7	2.3	virginica

Visualisation