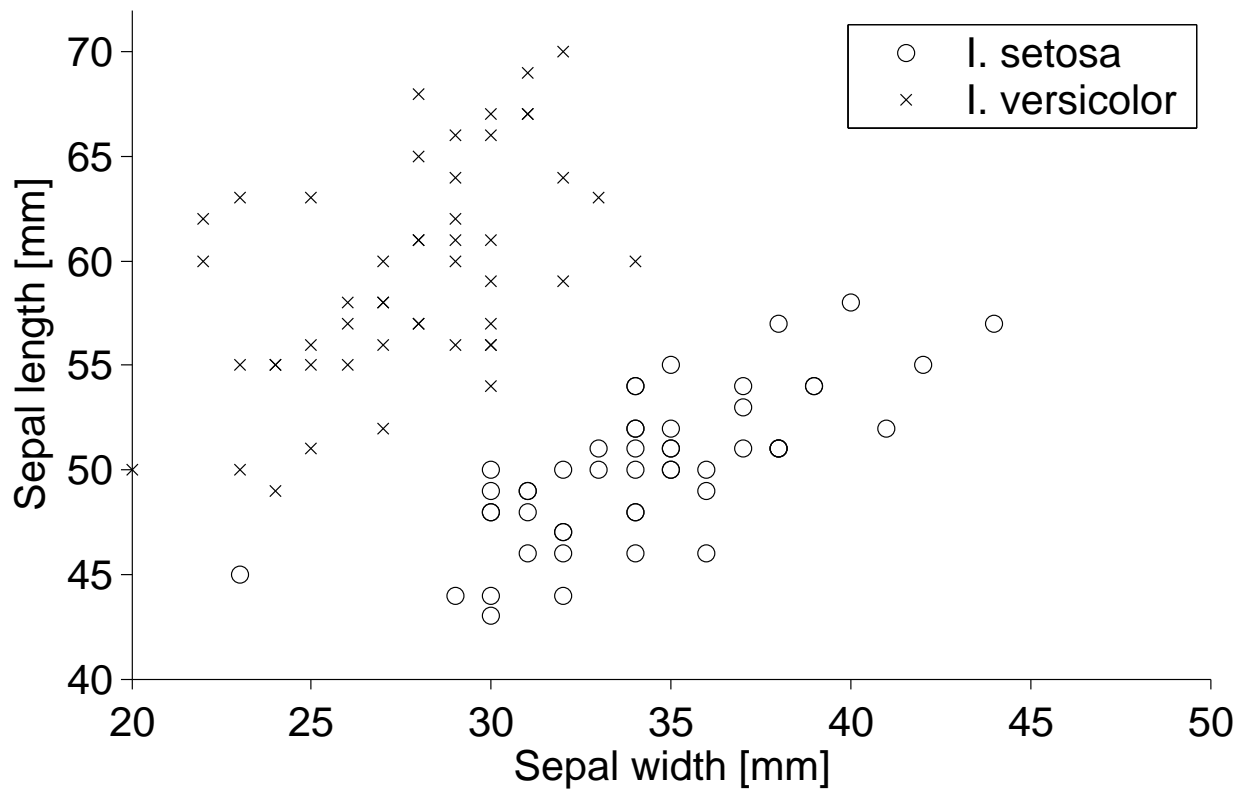


Fisher's (1936) Iris Data



Preliminaries of Discriminant Analysis

- Length measurements are positive, so we transform data by taking logarithms
- Regression of dummy variable ($y = 1$ for I. setosa and $y = 0$ for I. vericolor) against logarithms of sepal width and sepal length yields $R^2 = 0.84$ (significant at $p < 10^{-3}$)
⇒ Significant differences between groups (evident from plot)

Discriminant Analysis

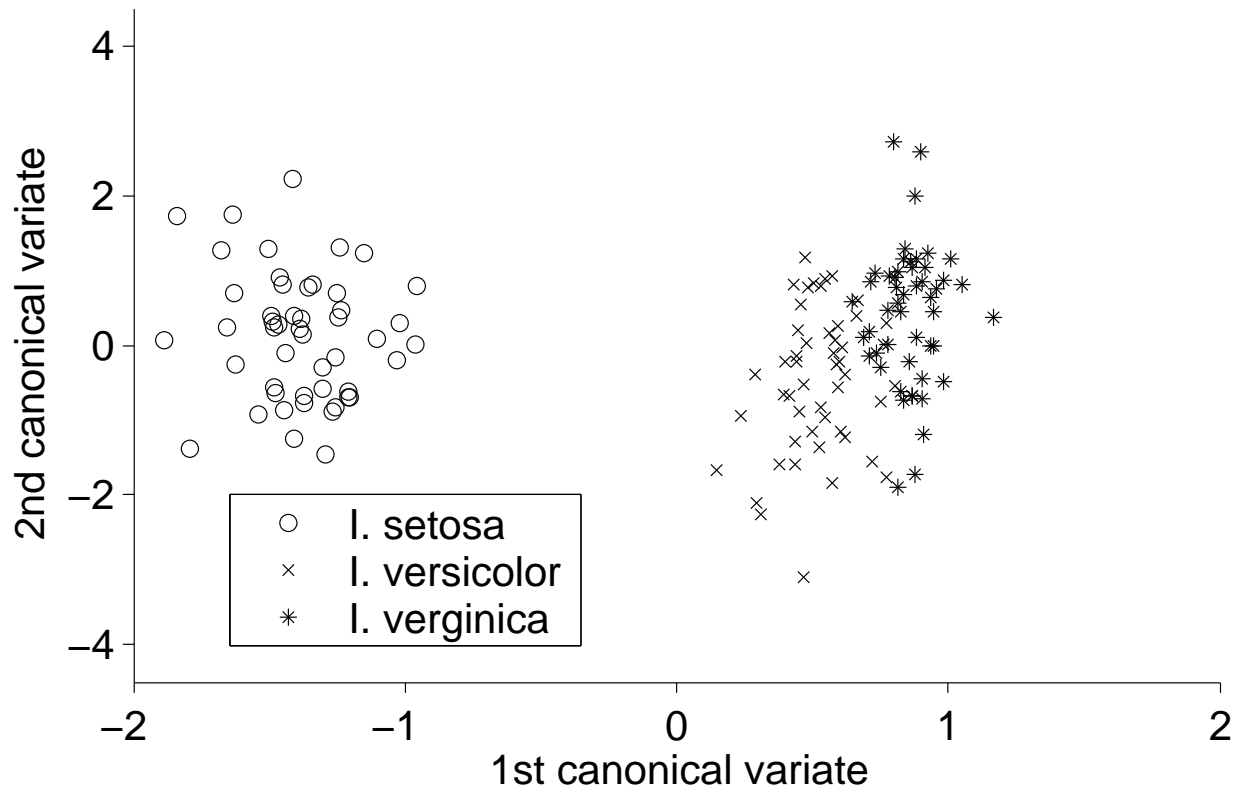
Discriminant analysis of transformed variables
[$x_1 = \log(\text{sepal width})$, $x_2 = \log(\text{sepal length})$]
yields weight vector

$$u_1 = (-3.9, 5.8)^T$$

and canonical variate

$$z_1 = Xu_1.$$

Scatterplot canonical variates



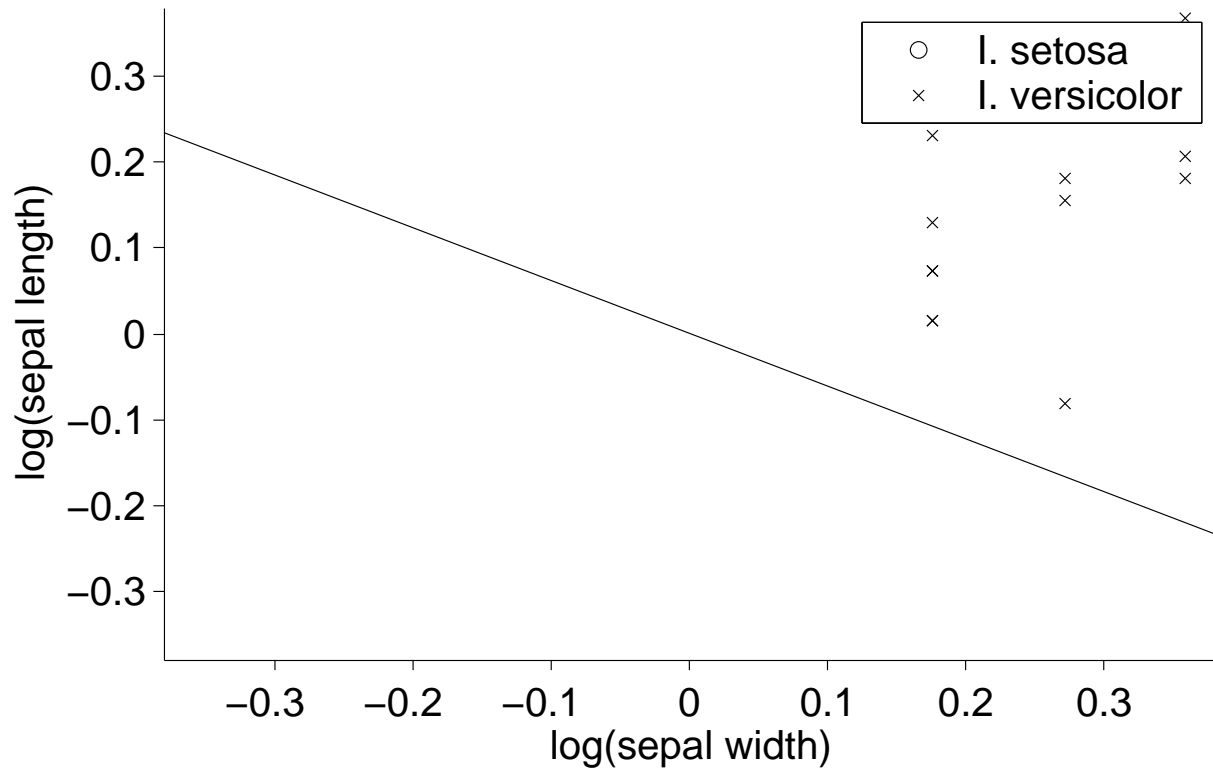
Classification

Allocate an observation x ($1 \times p$ vector) to the group i for which

$$\|(x - \mu_i)u_1\|^2,$$

the distance from group mean μ_i projected onto canonical variate, is smallest.

Decision Boundary



(for centered and log-transformed data)