



Module 5: Final Activity + Review Quiz

- Look at a real-world problem and choose an ML approach
- Final quiz (via Kahoot)
- Class discussion and wrap-up

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-
Mission 4 “Education and Research” - Component 2: “From research to business” - Investment
3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”



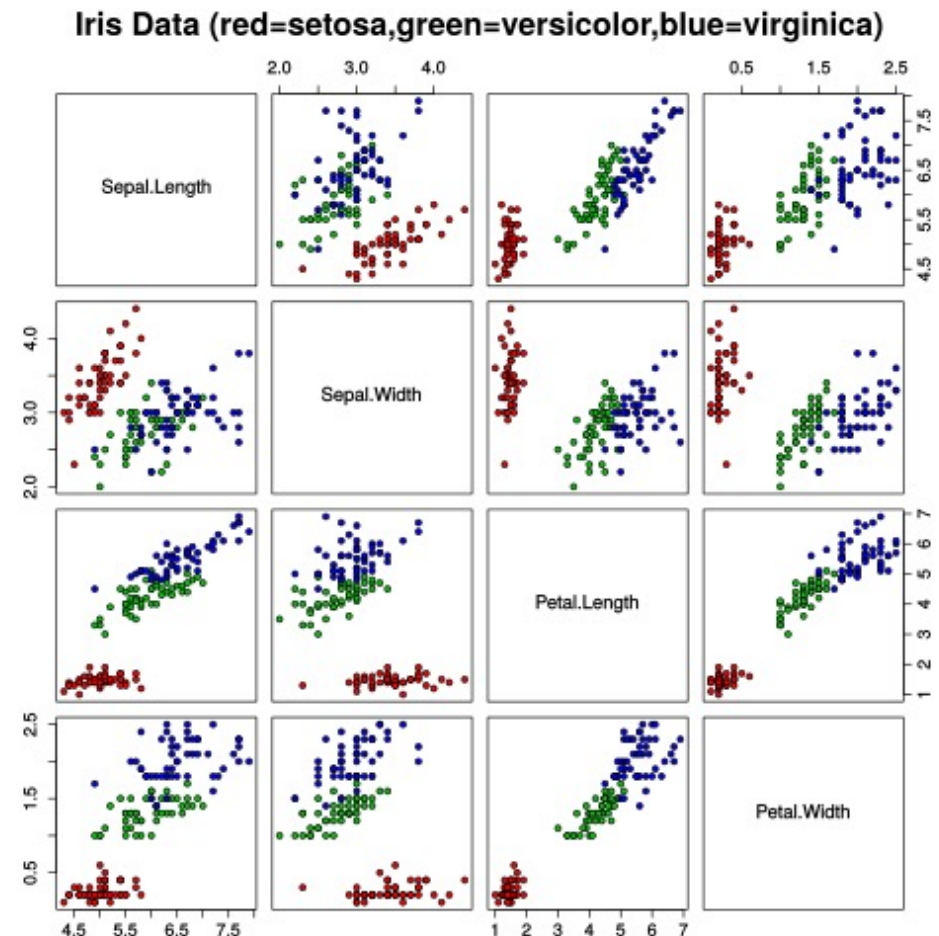
Iris flower data set (Fisher's Iris by the British statistician)

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 centimetres.

Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish each species.


Fisher's paper was published in the *Annals of Eugenics* (today the *Annals of Human Genetics*).





THANKS!

Francesco Iarlori

 <https://uk.linkedin.com/in/thefrankie/>

 @thefrankie

 Francesco@iarlori.com

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-
Mission 4 "Education and Research" - Component 2: "From research to business" - Investment
3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
INNOVAZIONE
DIGITALE





BackUp

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-
Mission 4 “Education and Research” - Component 2: “From research to business” - Investment
3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

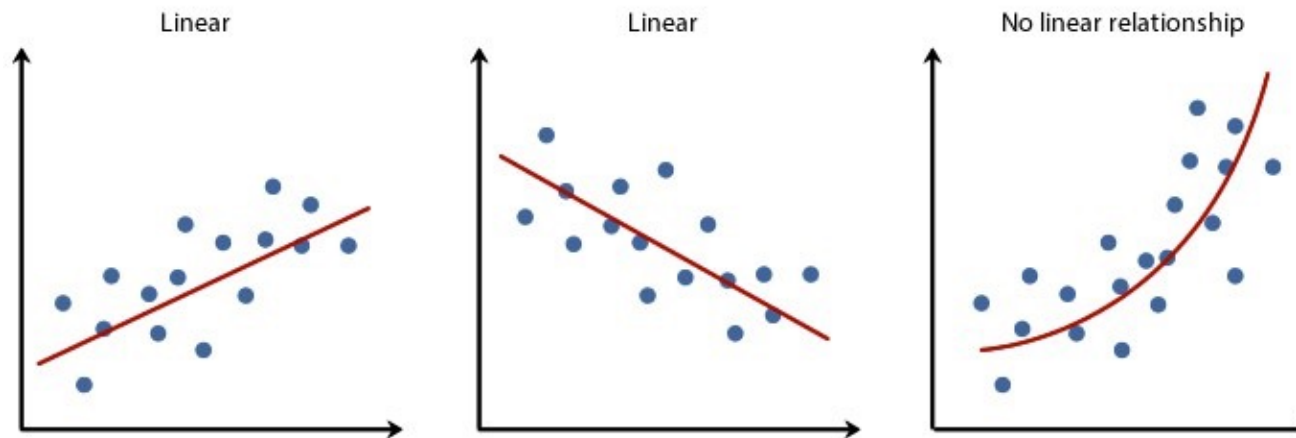


Word embeddings

- 🌐 Idea: learn an embedding from words into vectors
- 🌐 Need to have a function $W(\text{word})$ that returns a vector encoding that word.

ML problems (cont.)

- 🌐 **Supervised.** Provide samples of data and labels.
- 🌐 **Regression.** Estimating the relationships between a dependent variable and one or more independent variables



ML problems (cont.)

- 🌐 **Unsupervised.** Provide data items and group them
- 🌐 Clustering of objects in groups by similarity

