

1. (Esercizio del Laboratorio precedente) Data la matrice di dati della Tabella 11.5 (T11-5.DAT in JW) si considerino le tre famiglie di IRIS (setosa, versicolour, virginica) con le seguenti variabili:

Lung. Sepalo	Larg. Sepalo	Lung. Petalò	Larg. Petalò	Famiglia
5.1	3.5	1.4	0.2	1
4.9	3.0	1.4	0.2	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

- (a) Fare il diagramma di dispersione (`plot`) per le variabili  $(x_1, x_3)$  (lunghezza sepalo e lunghezza petalo) delle famiglie *setosa* e *virginica* (i due diagrammi sullo stesso grafico, con valori degli assi e legenda).
- (b) Per le osservazioni di ogni popolazione, determinare i vettori media (funzione `mean` di Matlab), le matrici di covarianza (funzione `cov` di Matlab) e le matrici di correlazione (funzione `corrcoef` di Matlab). Commentare i valori delle 3 matrici di correlazione.
- (c) Per le tre famiglie, considerare le variabili “di forma”:  
 $y_1$ : rapporto lunghezza sepalo e larghezza sepalo  
 $y_2$ : rapporto lunghezza petalo e larghezza petalo  
Fare un diagramma di dispersione per le tre famiglie e commentare i risultati. Calcolare e commentare le matrici di correlazione, rispetto ai risultati precedenti.
2. La matrice di dati ACID RAIN (da DASL and MORE DATA sul sito) riporta la distribuzione di precipitazioni acide in 47 siti negli UK; si veda anche la descrizione nel sito (attenzione al dato NA!)  
Fare un’analisi di clustering (funzione `clustering.m` nel sito) sia per i tipi di variabili che per i siti, e commentare i risultati.