

Laboratorio di Apprendimento Automatico

Fabio Aioli

Università di Padova

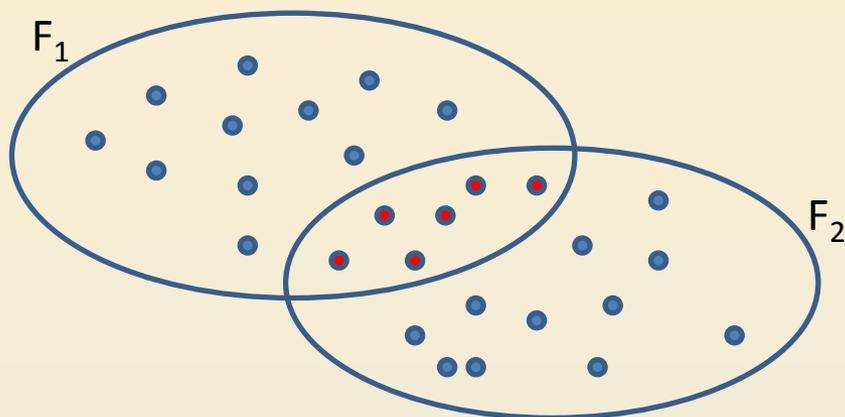
Laboratorio 20/11/11

- Introduzione a WEKA
 - Installazione, GUI, Explorer
 - Tab Preprocessing e Visualize
- Concetti
 - Relazioni (relations)
 - Attributi (attributes)
 - Valori
- File arff e csv

Laboratorio 20/11/11

- Analisi delle features..
- Per ogni feature (att:val)
 - Frequenza, ovvero il numero di volte che compare un determinato valore per un determinato attributo
- Per ogni coppia (att1:val1) (att2:val2)
 - Frequenza congiunzione, ovvero quante volte due coppie attributo-valore compaiono insieme negli esempi
 - Frequenza disgiunzione, ovvero quante volte almeno una coppia attributo-valore compare negli esempi

Similarità tra features



F_1 = insieme di esempi che hanno una certa coppia (att1:val1)

F_2 = insieme di esempi che hanno una certa coppia (att2:val2)

$ F_1 \cap F_2 $	Intersezione	$\frac{ F_1 \cap F_2 }{ F_1 \cup F_2 }$	Jaccard
$\frac{2 F_1 \cap F_2 }{ F_1 + F_2 }$	Dice	$\frac{ F_1 \cap F_2 }{\min(F_1 , F_2)}$	Overlap Coefficient

Esercizio

Dato il dataset *glass-lenses* studiato in laboratorio e utilizzando il programma *feat-analyze.cpp* calcolare quale tra le features è la più correlata con la classe e se questa caratteristica dipende o meno dal tipo di similarità utilizzata.