

Riconoscimento e Recupero dell'Informazione per Bioinformatica

LAB. 7 – PRTools (1)

Pietro Lovato

Corso di Laurea in Bioinformatica
Dip. di Informatica – Università di Verona
A.A. 2016/2017

PRTools

PRTools è un toolbox esterno a Matlab contenente funzioni per la costruzione di sistemi di Pattern Recognition.

PRTools include procedure per la generazione dei dati, addestramento di classificatori, selezione ed estrazione di feature, validazione.

PRTools

- Scaricabile gratuitamente da <http://www.37steps.com/software/> (PRTools5 !)
- Una volta fatto, è necessario aggiungere al path di matlab la cartella di prtools e le relative sottocartelle
- Si può fare da interfaccia grafica con il comando

```
>> pathtool
```

- Cliccare su “Add with subfolders”

Strutture dati fondamentali

- **prdataset** → è un tipo di variabile che memorizza:
 - Le rappresentazioni vettoriali degli oggetti del problema
 - Le etichette di classe
 - I prior sulle classi
 - Altre annotazioni sugli oggetti, le feature, il dataset in generale

Dataset: in pratica

```
>> D = prdataset(data, labels)
```

- data: matrice di dimensioni [n x k]
 - n oggetti
 - k feature
- labels: vettore colonna [n x 1]

Esercizio 1

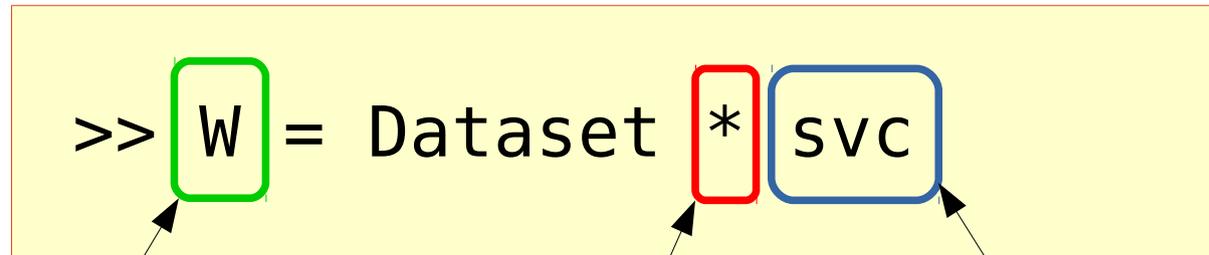
- Seguire il tutorial riportato al seguente indirizzo:

<http://www.37steps.com/prtools/elements/datasets/>

Strutture dati fondamentali

- **mapping** → funzione che definisce come trasformare i dati:
 - Normalizzazioni di misure grezze
 - Selezione/estrazione di feature
 - **Addestrare un classificatore**
 - Stima della likelihood
 - Trasformazione nello spazio di output di un classificatore (es. posterior, distanze, confidenze, ...)
 - Scelta della classe più probabile

Mapping: un semplice esempio



Mapping
addestrato

Applicazione del
mapping al dataset:
Sto addestrando un
classificatore, non
facendo una
moltiplicazione fra
matrici!

Mapping non
addestrato:
Funzione generica
che definisce il
classificatore

Mapping: in pratica

- Fase 1: costruzione del mapping su un dataset di train

```
>> W = D_train * svc %support vector machine
```

- Fase 2: applicazione del mapping ad un dataset di test

```
>> A = D_test * W
```

Fase 3: valuto la bontà del classificatore

```
>> testc(A)
```

Altri comandi utili

- Per visualizzare un dataset (max. 2 feature)

```
>> scatterd(D)
```

- Per plottare il confine di decisione di un classificatore

```
>> W = D*svc
```

```
>> plotc(W)
```

Esercizio 2

Seguire i tutorial riportati ai seguenti indirizzi:

- <http://www.37steps.com/prtools/examples/>
- <http://www.37steps.com/prtools/elements/mappings/#examples>
- <http://www.37steps.com/prtools/examples/classifiers/>

Esercizio 3

Una volta che si è presa confidenza con i comandi di PRTools, caricare in memoria il file `dataset.mat`, ed eseguire i seguenti esercizi.

- Costruire un `prdataset` per i dati di train e uno per quelli di test
- Plottare con il comando `scatterd` i dati di train

Esercizio 3

- Addestrare il classificatore di prtools knnc sui dati di train
- Plottare il confine di decisione del classificatore con il comando `plotc`
- Testare l'errore di classificazione sui dati di test (funzione `testc`)

Esercizio 3

- Addestrare il classificatore di prtools svc sui dati di train
 - Provare a variare il parametro C e il tipo di kernel:
 - Valori ragionevoli per C = [0.01 0.1 1 10 100];
 - Provare con il kernel lineare e rbf

```
>> D*svc() % kernel lineare, C=1
```

```
>> D*svc([], 10) % kernel lineare, C=10
```

```
>> D*svc([], 'r', [], 10) % kernel rbf, C=10
```

Esercizio 3

- Plottare i confini di decisione di knnc e di svc
- Quale classificatore funziona meglio? Perché?

