

## CLASSIFICATORI

- Implementazione di un classificatore lineare
- Classificatore quadratico nei prttools

1

## IMPLEMENTAZIONE DI UN CLASSIFICATORE LINEARE

**Modello Gaussiano: caso  $\Sigma_i = \sigma^2 \mathbf{I}$**

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0}$$

$$w_i = \frac{1}{\sigma^2} \mu_i; w_{i0} = \frac{1}{2\sigma^2} \mu_i^T \mu_i + \ln(P(\omega_i))$$

Vedere il file linearDiscriminant.m

function label=linearDiscriminant(trainSet,labelTrainSet, testSet);

2

## CLASSIFICATORE QDC

QDC Quadratic Bayes Normal Classifier (Bayes-Normal-2)

$$\mathbf{W} = \text{QDC}(\mathbf{A}, \mathbf{R}, \mathbf{S})$$

### INPUT

A Dataset

R,S Regularization parameters,  $0 \leq R, S \leq 1$   
(optional; default: no regularization, i.e.  $R, S = 0$ )

### DESCRIPTION

Computation of the quadratic classifier between the classes of the dataset A assuming normal densities. R and S ( $0 \leq R, S \leq 1$ ) are regularization parameters used for finding the covariance matrix

3

Testate il classificatore QDC col dataset artificiale ottenuto da GENDATL Generation of Lithuanian classes

- generate il dataset
- dividete in train e test
- addestrate il classificatore sul train
- calcolate l'errore sul train e sul test
- confrontate i risultati con quelli ottenuti col classificatore LDC
- visualizzate entrambe le superfici di decisione

	Err Train	Err Test
LDC	25%	18.06%
QDC	12.50%	9.72%

Considerazioni:

- quanto influisce la scelta casuale dei dati?
- quanto influisce la scelta dei parametri?
- quanto influisce la dimensione del training set?

4

```

A= gendatl(300);
[ds_train, ds_test]=gendat(A, 0.10);
lineare= ldc(ds_train);
quad= qdc(ds_train);
errTrainLin= testc(ds_train*lineare);
errTrainQuad= testc(ds_train*quad);
errTestLin= testc(ds_test*lineare);
errTestQuad= testc(ds_test*quad);

```

```

scatterd(ds_test)
hold on
plotc(lineare)
plotc(quad)
matriceErrore=[errTrainLin, errTestLin;
errTrainQuad,errTestQuad]*100;
openvar('matriceErrore')

```

5

## STIMARE L'ERRORE MEDIANTE CROSS-VALIDATION

Error estimation by cross validation

[ERR,CERR,NLAB\_OUT] = CROSSVAL(A,CLASSF,N,1,FID)

### INPUT

A Input dataset  
CLASSF The untrained classifier to be tested.  
N Number of dataset divisions  
(default: N==number of samples, leave-one-out)  
FID File descriptor for progress report file (default: 0)

### OUTPUT

ERR Average test error weighted by class priors.  
CERR Unweighted test errors per class  
NLAB\_OUT Assigned numeric labels

6

[ERR,STDS] = CROSSVAL(A,CLASSF,N,NREP,FID)

### INPUT

A Input dataset  
CLASSF The untrained classifier to be tested.  
N Number of dataset divisions  
(default: N==number of samples, leave-one-out)  
NREP Number of repetitions (default: 1)  
FID File descriptor for progress report file (default: 0)

### OUTPUT

ERR Average test error weighted by class priors.  
STDS Standard deviation over the repetitions.

7

## ESEMPIO

Generate un dataset e valutate le prestazioni di LDC e QDC con 5 fold, con e senza ripetizioni

Traccia:

w=qdc;

[ERR,CERR]=crossval(A,w,5,1)

[ERR,STDS] = crossval (A,w,5,20)

8