Pattern Recognition
and Applications Lab

Elements of Linear Discriminant Functions
Exercises

Battista Biggio

battista.biggio@diee.unica.it

Dept. Of Electrical and Electronic Engineering
University of Cagliari, Italy

University
of Cagliari, Italy

Department of
Electrical and Electronic
Engineering

Exercise 1

- Consider the data

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ -1 & -2 \\ -2 & -1 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

- and the linear classification function $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$, with $\boldsymbol{w} = [1\ 1]^T$ and $b = 0$.

- Points are classified as positive if $f(\boldsymbol{x}) \geq 0$

- Classify the data points in $\mathbf{X}$ and compute the overall classification error

Exercise 1: Solution

- Samples in $\mathbf{X}$ are stored as rows, hence it is not difficult to compute the values of the discriminant function as

$$f(\mathbf{X}) = \boldsymbol{X}\boldsymbol{w} + b = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ -1 & -2 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ -3 \\ -3 \end{bmatrix}$$

- which, compared against 0, gives the classification labels

$$\boldsymbol{y}_c = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

- Since they are exactly equal to the true labels $\boldsymbol{y}$, the classification error is zero

Exercise 2

- Given the following data samples

$$\mathbf{X} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 2 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix},$$

- find the linear discriminant function using the **Batch perceptron algorithm** and the criterion function

$$L_p(\boldsymbol{w}, b) = -\sum_{i:y_i f(\boldsymbol{x}_i)<0} y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b)$$

- Initialize $\boldsymbol{w} = [0.1, 0.1]^T,\ b = 0.1,\ \eta = 1,\ \theta = 0$
- Use the L_1 norm to compute the termination condition

Exercise 2: Solution

```
begin initialize w, b, η, θ
    do
      select samples for which y_i f(x_i) < 0
      update w ← w − η∇_w L_p(w, b) and b ← b − η∇_b L_p(w, b)
          (using the selected samples)
    until η(|∇_w L_p(w, b)| + |∇_b L_p(w, b)|) < θ
end
```

- In the lecture slides the algorithm is the same, but uses the augmented vector **w**, hence b is implicitly considered in **w**

- Here we have to compute the derivatives w.r.t $\boldsymbol{w}$ and b:

$$\nabla_{\boldsymbol{w}} L_p(\boldsymbol{w}, b) = -\sum_{i: y_i f(\boldsymbol{x}_i) < 0} y_i \boldsymbol{x}_i$$

$$\nabla_b L_p(\boldsymbol{w}, b) = -\sum_{i: y_i f(\boldsymbol{x}_i) < 0} y_i$$

Exercise 2: Solution

- Recall that we are considering a **batch** algorithm
  - all samples are considered for updating ***w*** and b

Iteration 1

- We first have to classify the data points with the current ***w*** and b

$$f(\mathbf{X}) = \boldsymbol{Xw} + b = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 2 \\ 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} + 0.1 = \begin{bmatrix} 0 \\ 0 \\ 0.1 \\ 0.3 \\ 0.4 \\ 0.4 \end{bmatrix}$$

Exercise 2: Solution

- Then, consider the classification errors for which $y_i f(\boldsymbol{x}_i) < 0$

$$\boldsymbol{y} f(\mathbf{X}) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \circ \begin{bmatrix} 0 \\ 0 \\ 0.1 \\ 0.3 \\ 0.4 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0.1 \\ -0.3 \\ -0.4 \\ -0.4 \end{bmatrix}$$

- The three last patterns are wrongly classified, so they will be used to update ***w*** and b

$$\nabla_{\boldsymbol{w}} L_p(\boldsymbol{w}, b) = -\sum_{i: y_i f(\boldsymbol{x}_i) < 0} y_i \boldsymbol{x}_i = +1 \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

$$\nabla_b L_p(\boldsymbol{w}, b) = -\sum_{i: y_i f(\boldsymbol{x}_i) < 0} y_i = 3$$

Exercise 2: Solution

- **Parameter updates**
  $\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \nabla_{\boldsymbol{w}} L_p(\boldsymbol{w}, b)$ and $b \leftarrow b - \eta \nabla_b L_p(\boldsymbol{w}, b)$

$$\boldsymbol{w} = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} - 1 \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} -2.9 \\ -4.9 \end{bmatrix}$$

$$b = 0.1 - 1 * 3 = -2.9$$

- **Termination condition** $\eta(|\nabla_{\boldsymbol{w}} L_p(\boldsymbol{w}, b)| + |\nabla_b L_p(\boldsymbol{w}, b)|) < \theta$

$$t = 1 * (|3| + |5| + |3|) < \theta = 0$$

- Termination condition is *False* (then, loop continues)

Exercise 2: Solution

iter	$\boldsymbol{y}f(\boldsymbol{x})$	L_p	$\nabla_{\boldsymbol{w}}L_p, \nabla_b L_p$	$\boldsymbol{w}, b$	t	θ
1.	[0. 0. 0.1 -0.3 -0.4 -0.4]	1.1	[3 5] 3	[-2.9 -4.9] -2.9	11.0	0
2.	[0. 2. -2.9 12.7 15.6 13.6]	2.9	[0 0] -1	[-2.9 -4.9] -1.9	1.0	0
3.	[1. 3. -1.9 11.7 14.6 12.6]	1.9	[0 0] -1	[-2.9 -4.9] -0.9	1.0	0
4.	[2. 4. -0.9 10.7 13.6 11.6]	0.9	[0 0] -1	[-2.9 -4.9] 0.1	1.0	0
5.	[3. 5. 0.1 9.7 12.6 10.6]	0.0	[0 0] 0	[-2.9 -4.9] 0.1	0.0	0

Homework: Plot the data points and the decision function at each iteration

Exercise 3

- Given a regression problem and the following data samples

$$\mathbf{X} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 2 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{bmatrix},$$

- find the linear discriminant function using the **Widrow-Hoff** (batch) **algorithm,** which minimizes

$$L_r(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{w}^T\boldsymbol{x}_i + b - y_i)^2$$

- Initialize $\boldsymbol{w} = [0.1, 0.1]^T,\ b = 0.1,\ \eta = 0.1,\ \theta = 0.06$

Exercise 3: Solution

- The Widrow-Hoff algorithm (batch version) is

begin initialize w, θ, η, $k=0$

repeat

$\boldsymbol{w}=\boldsymbol{w} - \eta\, \nabla_{\boldsymbol{w}} L_r(\boldsymbol{w}, b)$

$b=b - \eta\, \nabla_b L_r(\boldsymbol{w}, b)$

until $\eta\,(|\nabla_w L_r(\boldsymbol{w}, b)| + |\nabla_b L_r(\boldsymbol{w}, b)|) < \theta$

- The derivatives w.r.t $\boldsymbol{w}$ and b:

$$\nabla_{\boldsymbol{w}} L_r(\boldsymbol{w}, b) = \sum_{i=1}^{N} (\boldsymbol{w}^T \boldsymbol{x}_i + b - y_i)\, \boldsymbol{x}_i$$

$$\nabla_b L_r(\boldsymbol{w}, b) = \sum_{i=1}^{N} (\boldsymbol{w}^T \boldsymbol{x}_i + b - y_i)$$

Exercise 3: Solution

θ=0.06

iter	$f(\boldsymbol{x})$	L_r	$\nabla_{\boldsymbol{w}} L_{\mathrm{r}}, \nabla_b L_r$	$\boldsymbol{w}, b$	t
1.	0.1*[0 0 1 3 4 4]	0.24	[1. 1.4], 0.6	[0. -0.04], 0.04	0.3
2.	0.01*[4 8 4 -4 -4 0]	0.057	[-0.28 -0.64], -0.52	[0.028 0.024], 0.092	0.14
3.	0.01*[6 7 9 14 17 17]	0.014	[0.25 0.32], 0.104	[0.0032 -0.008], 0.0816	0.07
4.	0.01*[8 9 8 7 7 8]	0.003	[-0.05 -0.14], -0.136	[0.008 0.006], 0.09	0.03

Homework: Plot the data points and the decision function at each iteration