Intelligent Systems
Exercises Block 2 Chapter 2
Learning discriminant functions: Perceptron algorithm

Escuela Técnica Superior de Ingeniería Informática
Dep. de Sistemas Informáticos i Computación
Universitat Politècnica de València

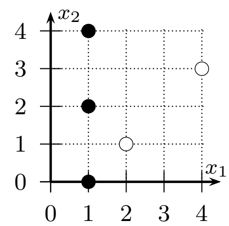
27 de octubre de 2016

Questions

1 $\boxed{\text{A}}$ The perceptron algorithm is a ...

A) supervised and linear classifier
B) supervised and non-linear classifier
C) non-supervised and linear classifier
D) non-supervised and non-linear classifier

2 $\boxed{\text{C}}$ The figure on the right represents five two-dimension training samples that belong to two classes ∘ and •. We want to build a linear classifier for any $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, as follows:



$$c(\mathbf{x}) = \begin{cases} \circ & \text{if } \mathbf{w}^t\,\mathbf{x} > 0 \\ \bullet & \text{if } \mathbf{w}^t\,\mathbf{x} \leq 0 \end{cases} \quad \text{where} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \text{ is a weight vector to be selected}$$

If our learning criterion is the minimum classification error (over the learning samples), we will choose ...:

A) $\mathbf{w} = (1, 0)^t$
B) $\mathbf{w} = (1, 1)^t$
C) $\mathbf{w} = (1, -1)^t$
D) None of the above because we can find other weight vectors that produce a lower number of errors on the given training samples (there are better choices).

3 $\boxed{\text{D}}$ In the perceptron algorithm:

A) there are mainly two parameters: the number of classes and the number of prototypes
B) the learning step α must be as large as possible in order to learn as much as possible
C) the margin must be zero when the classes are no linear-separable
D) there are mainly two parameters: the learning step α and the margin b.

4 $\boxed{\text{B}}$ In the perceptron algorithm:

A) there are mainly two parameters: the number of classes and the number of prototypes
B) the margin b allows for finding adequate solutions when the problem is non-linearly separable
C) the margin b depends on the learning rate α
D) there are mainly two parametes: the learning rate α and the number of iterations

5 $\boxed{\text{B}}$ The parameter of the Perceptron algorithm that we name *margin*, b, is a real value that, assuming it is positive (as it usually is), reduces he number of possible solutions that the algorithm can find. Concretely, given N training samples, $(\mathbf{x}_1, c_1), \ldots, (\mathbf{x}_N, c_N)$ from C classes, the perceptron algorithm will find linear discriminant functions $g_1(\cdot), \ldots, g_C(\cdot)$ such that for every $n = 1, \ldots, N$:

A) $g_{c_n}(\mathbf{x}_n) > g_c(\mathbf{x}_n)$ for every class $c \neq c_n$
B) $g_{c_n}(\mathbf{x}_n) > g_c(\mathbf{x}_n) + b$ for every class $c \neq c_n$
C) $g_{c_n}(\mathbf{x}_n) > g_c(\mathbf{x}_n) - b$ for every class $c \neq c_n$
D) None of the above

6 $\boxed{\text{B}}$ Given a classification problem of C classes $C \in \{1, 2, \ldots, C\}$, where objects are represented with a feature vector of D dimensions, $\mathbf{x} \in \mathbb{R}^D$, and assuming that a given $\mathbf{x}$ belongs to class 1, the Perceptron algorithm:

A) Modifies the linear discriminant $g_1(\mathbf{x})$ in any case.
B) Modifies the linear discriminant $g_1(\mathbf{x})$ if it exists $c \neq 1, g_c(\mathbf{x}) > g_1(\mathbf{x})$.
C) Modifies the linear discriminant $g_c(\mathbf{x})$ if $g_c(\mathbf{x}) < g_1(\mathbf{x})$ with $c \neq 1$.
D) Modifies the linear discriminant $g_1(\mathbf{x})$ only if $g_c(\mathbf{x}) > g_1(\mathbf{x})$ for every $c \neq 1$.

7 $\boxed{\text{C}}$ Given a classification problem of two classes, the following two-dimensional samples are provided: $\mathbf{x}_1 = (1,1)^t, \mathbf{x}_2 = (2,2)^t, \mathbf{x}_3 = (2,0)^t$; $\mathbf{x}_1$ and $\mathbf{x}_2$ belong to class A and $\mathbf{x}_3$ to class B. Taking into account that we are using a classifier based on linear discriminant functions with weight vectors $\mathbf{w}_A$ and $\mathbf{w}_B$ corresponding to classes A and B respectively, which of the following statements is *false*:

A) It is possible to find a linear discriminant function that classifies $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ with error=2/3.
B) Weight vectors $\mathbf{w}_A = (1, -1, 1)^t$ and $\mathbf{w}_B = (1, 2, -4)^t$ classify $\mathbf{x}_1$, $\mathbf{x}_2$ y $\mathbf{x}_3$ without errors.
C) Weight vectors $\mathbf{w}_A = (1, -1, 1)^t$ and $\mathbf{w}_B = (1, 2, -4)^t$ classify $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ with error=1/3.
D) It is possible to find a discriminant function that classifies $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ with error=1/3.

8 $\boxed{\text{D}}$ Let be a classification problem of three classes $\{A, B, C\}$ where objects are represented in a two-dimensional space $\mathbb{R}^2$. We want to use a classifier based on linear discriminant functions with the following weight vector for each class, $\mathbf{w}_A = (1, 1, 0)^t$, $\mathbf{w}_B = (-1, 1, -1)^t$ and $\mathbf{w}_C = (1, -2, 2)^t$. Which is the classification of $\mathbf{x}_1 = (1, 1)^t$ and $\mathbf{x}_2 = (0, -1)^t$?

A) $c(\mathbf{x}_1) = B \quad c(\mathbf{x}_2) = C$
B) $c(\mathbf{x}_1) = A \quad c(\mathbf{x}_2) = B$
C) $c(\mathbf{x}_1) = B \quad c(\mathbf{x}_2) = A$
D) $c(\mathbf{x}_1) = A \quad c(\mathbf{x}_2) = A$

9 $\boxed{\text{D}}$ (Exam 18th January 2013) Let $g_1(\mathbf{y}) = y_1^2 + 2\,y_2^2$ and $g_2(\mathbf{y}) = 2\,y_1^2 + y_2^2$ be two discriminant functions for classes 1 and 2, respectively. The decision boundary between these two classes is:

A) A parabola
B) Hyperspherical
C) It is given by the equation $y_1^2 + y_2^2 = 0$.
D) A straight line: $y_2 = y_1$ – (by doing $g_1(\mathbf{y}) = g_2(\mathbf{y})$)

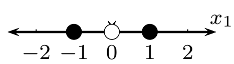
10 $\boxed{\text{A}}$ (Exam 30th January 2013) For a two-class classification problem in $\Re^2$ we have three different classifiers. One is formed by the two linear discriminant functions: $g_1(y) = 3 + 4\,y_1 - 2\,y_2$ and $g_2(y) = -3 + 1.5\,y_1 + 5\,y_2$. The second classifier is formed by $g_1'(y) = 6 + 8\,y_1 - 4\,y_2$ and $g_2'(y) = -6 + 3\,y_1 + 10\,y_2$. And the third by $g_1''(y) = -6 - 8\,y_1 + 4\,y_2$ and $g_2''(y) = 6 - 3\,y_1 - 10\,y_2$. Are the three classifiers equivalent? That is, do they define the same decision boundaries?

A) (g_1, g_2) y (g_1', g_2') are equivalent.
B) The three of them are equivalent.
C) (g_1, g_2) y (g_1'', g_2'') are equivalent.
D) (g_1', g_2') y (g_1'', g_2'') are equivalent.

11 $\boxed{\text{A}}$ (Exam 30th January 2013) The perceptron algorithm is a ...

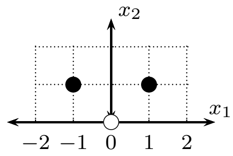
A) supervised and lineal classifier
B) supervised and quadratic classifier
C) non-supervised and linear classifier
D) non-supervised and quadratic classifier

12 $\boxed{\text{B}}$ (Exam January 15, 2014) The figure on the right shows three one-dimensional samples classified in two classes ○ and ●. Which is the number of errors of a minimum-error linear classifier?



A) 0
B) 1
C) 2
D) 3

13 $\boxed{\text{A}}$ (Exam January 15, 2014) Consider that we add a new feature x_2 to the samples of the above question. x_2 is defined as $x_2 = x_1^2$. Thus, we have now three two-dimensional samples as can be observed in the figure on the right. In this case, which is the number of errors of a minimum-error linear classifier?



A) 0
B) 1
C) 2
D) 3

14 $\boxed{\text{A}}$ (Exam January 15, 2014) Consider a classification problem in 2 classes, $c = 1, 2$, for objects represented by means of two-dimensional feature vectors. We have two training samples: $\mathbf{x}_1 = (0, 0)^t$ belongs to class $c_1 = 1$, and $\mathbf{x}_2 = (1, 1)^t$ belongs to class $c_2 = 2$. Likewise, we have the following linear classifier: $\mathbf{w}_1 = (w_{10}, w_{11}, w_{12}) = (1, -1, -1)^t$ and $\mathbf{w}_2 = (w_{20}, w_{21}, w_{22}) = (-1, 1, 1)^t$. If we apply one iteration of the Perceptron algorithm with learning speed $\alpha = 1$ and margin $b = 0.1$, then:

A) None of the weight vectors will be modified.
B) The weight vector of class 1 will be modified.
C) The weight vector of class 2 will be modified.
D) Both weight vectors will be modified.

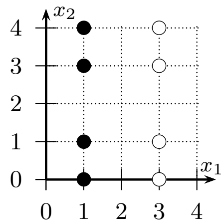
15 $\boxed{\text{C}}$ (Exam January 15, 2014) The Perceptron algorithm is controlled by two parameters, the *learning speed*, α, and the *margin*, b, both real values. Assuming we don't know whether the training samples are linearly separable or not, which values of the parameters α and b provide decision boundaries of better quality?

A) $\alpha = 0.1$ and $b = 0.0$.
B) $\alpha = 0.0$ and $b = 0.0$.
C) $\alpha = 0.1$ and $b = 1.0$.
D) $\alpha = 0.0$ and $b = 1.0$.

16 $\boxed{\text{B}}$ (Exam January 28, 2014) Consider a classification problem in two classes, $c = A, B$, for objects represented by means of two-dimensional feature vectors. The weight vectors that result from applying the Perceptron algorithm with a training data set are $\mathbf{w}_A = (1, 1, 0)^t$ and $\mathbf{w}_B = (-1, 0, 1)^t$. Given these results, which class do samples $\mathbf{x}_1 = (-1, 0)^t$ and $\mathbf{x}_2 = (0, 3)^t$ belong to?

A) $\hat{c}(\mathbf{x}_1) = A$ and $\hat{c}(\mathbf{x}_2) = A$. $\quad \mathbf{x}_1 \;:\; \mathbf{w}_A^t \cdot (1, -1, 0)^t = 0 \quad \mathbf{w}_B^t \cdot (1, -1, 0)^t = -1 \;\Rightarrow\; \mathbf{x}_1 \in A$
B) $\hat{c}(\mathbf{x}_1) = A$ and $\hat{c}(\mathbf{x}_2) = B$.
C) $\hat{c}(\mathbf{x}_1) = B$ and $\hat{c}(\mathbf{x}_2) = A$. $\quad \mathbf{x}_2 \;:\; \mathbf{w}_A^t \cdot (1, 0, 3)^t = 1 \quad \mathbf{w}_B^t \cdot (1, 0, 3)^t = 2 \;\Rightarrow\; \mathbf{x}_2 \in B$
D) $\hat{c}(\mathbf{x}_1) = B$ and $\hat{c}(\mathbf{x}_2) = B$.

17 $\boxed{\text{D}}$ (January 13, 2015) The figure on the right shows four two-dimensional training samples of 2 classes: ○ or ●. Assuming our learning criteria is to minimize the number of misclassified samples, we will select as the weight vector of each class . . .



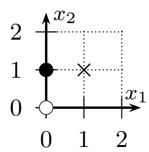
A) $\mathbf{a}_\circ = (3, 1, 1)^t$ y $\mathbf{a}_\bullet = (1, 2, 1)^t$ $\quad x_1 = 2 \quad R_\circ = \{x : x_1 < 2\}$ y $R_\bullet = \{x : x_1 > 2\}$

B) $\mathbf{a}_\circ = (1, 1, 2)^t$ y $\mathbf{a}_\bullet = (3, 1, 1)^t$ $\quad x_2 = 2 \quad R_\bullet = \{x : x_2 < 2\}$ y $R_\circ = \{x : x_2 > 2\}$

C) $\mathbf{a}_\circ = (3, 1, 1)^t$ y $\mathbf{a}_\bullet = (1, 1, 2)^t$ $\quad x_2 = 2 \quad R_\circ = \{x : x_2 < 2\}$ y $R_\bullet = \{x : x_2 > 2\}$

D) $\mathbf{a}_\circ = (1, 2, 1)^t$ y $\mathbf{a}_\bullet = (3, 1, 1)^t$ $\quad x_1 = 2 \quad R_\bullet = \{x : x_1 < 2\}$ y $R_\circ = \{x : x_1 > 2\}$

18 $\boxed{\text{B}}$ (January 13, 2015) The figure on the right shows three two-dimensional training samples of three classes: ○, ● and ×. Given the weight vectors $\mathbf{a}_\circ = (-2, -1, -3)^t$, $\mathbf{a}_\bullet = (-1, -3, 1)^t$ and $\mathbf{a}_\times = (-3, 3, -1)^t$, how many misclassification errors are generated?



A) 0
B) 1
C) 2
D) 3

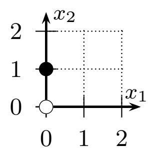
19 $\boxed{\text{A}}$ (January 13, 2015) If we apply one iteration of the Perceptron algorithm with learning rate $\alpha = 1.0$ and margin $b = 0.0$ over the samples and weight vectors of the above question, how many misclassification errors are generated with the new weight vectors?

A) 0 $\mathbf{a}_\circ = (-1,-1,-3)^t$, $\mathbf{a}_\bullet = (-2,-3,1)^t$ y $\mathbf{a}_\times = (-3,3,-1)^t$
B) 1
C) 2
D) 3

20 $\boxed{\text{C}}$ (January 26, 2015) Given a linear classifier of two classes $\circ$ and $\bullet$ with weight vectors $\mathbf{a}_\circ = (0,-1,1)^t$ and $\mathbf{a}_\bullet = (0,1,-1)^t$, which of the following weight vectors does ***NOT*** define an equivalent classifier to this one?

A) $\mathbf{a}_\circ = (1,-1,1)^t$ and $\mathbf{a}_\bullet = (1,1,-1)^t$ $f(z) = az + b$ with $a = 1$ and $b = 1$
B) $\mathbf{a}_\circ = (-1,-2,2)^t$ and $\mathbf{a}_\bullet = (-1,2,-2)^t$ $f(z) = az + b$ con $a = 2$ and $b = -1$
C) $\mathbf{a}_\circ = (0,2,-2)^t$ and $\mathbf{a}_\bullet = (0,-2,2)^t$ $f(z) = az + b$ with $a = -2$ and $b = 0$
D) $\mathbf{a}_\circ = (0,-2,2)^t$ and $\mathbf{a}_\bullet = (0,2,-2)^t$ $f(z) = az + b$ with $a = 2$ and $b = 0$

21 $\boxed{\text{A}}$ (January 26, 2015) The figure on the right shows two two-dimensional training samples of two classes: $\circ$, and $\bullet$. Given the weight vectors $\mathbf{a}_\circ = (0,1,-2)^t$, and $\mathbf{a}_\bullet = (0,0,1)^t$, if we apply one iteration of the Perceptron algorithm with learning rate $\alpha = 1.0$ and margin $b = 0.5$ over the training samples and weight vectors provided, how many misclassification errors will be generated with the new weight vectors resulting from the Perceptron algorithm?



A) 0 $\mathbf{a}_\circ = (1,1,-2)^t$ y $\mathbf{a}_\bullet = (-1,0,1)^t$
B) 1
C) 2
D) 3

22 $\boxed{\text{D}}$ (January, 2016) Let's consider a typical classification problem in C classes and objects represented through D-dimensional real feature vectors. In general, we can say that it it more difficult to find an accurate classifier when ...

A) the values of C and D are smaller
B) the value of C is smaller and the value of D is larger
C) the value of C is larger and the value of D is smaller
D) the values of C and D are larger

23 $\boxed{\text{B}}$ (January, 2016) We have learnt two different classifiers, c_A and c_B, for a classification problem. The probability of error of c_A has been empirically estimated from 100 test samples, obtaining an empirical estimate of error $\hat{p}_A = 0.10$ (10 %). Similarly, the probability of error of c_B has been empirically estimated but with a set of 200 test samples, obtaining an empirical estimate of error of 10 %, too ($\hat{p}_B = 0.10$). Based on these estimations, we can affirm with a 95 % of confidence that:

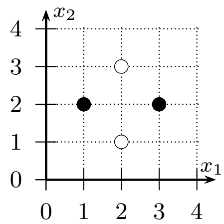
A) The confidence intervals of $\hat{p}_A$ y $\hat{p}_B$ are identical.

B) The confidence interval of $\hat{p}_A$ is larger than the one of $\hat{p}_B$. $I_A = \hat{p}_A \pm 1.96\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{100}} = 0.10 \pm 0.06$

C) The confidence interval of $\hat{p}_B$ is larger than the one of $\hat{p}_A$. $I_B = \hat{p}_B \pm 1.96\sqrt{\frac{\hat{p}_B(1-\hat{p}_B)}{200}} = 0.10 \pm 0.04$

D) In this case, the confidence intervals of $\hat{p}_A$ and $\hat{p}_B$ are irrelevant because the estimate of error is the same.

24 $\boxed{\text{C}}$ (January, 2016) The figure on the right shows 4 two-dimensional samples classified in 2 classes: $\circ$ and $\bullet$. If we apply the Perceptron algorithm with initial weight vectors $\mathbf{a}_\circ = (0,1,0)^t$ and $\mathbf{a}_\bullet = (0,0,1)^t$, a learning rate $\alpha > 0$ and a margin b, indicate which assertion is **CORRECT**:



A) The algorithm will converge for some $b > 0$
B) The algorithm only converges if $b \leq 0$
C) If $b > 0$ there is no convergence but, by adjusting α, we can obtain good solutions after a finite number of iterations with respect to the probability of classification error (with 25 % of misclassification error)
D) The algorithm is not applicable in this case because the classes are non-linearly separable.

25 $\boxed{\text{B}}$ (January, 2016) Which is the number of errors of a minimum-error linear classifier for the training samples of the above question?

A) 0
B) 1
C) 2
D) 3

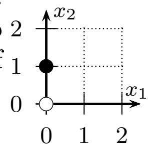
26 $\boxed{\text{B}}$ (January, 2016) Given a linear classifier of two classes ○ and ● with weight vectors $\mathbf{a}_\circ = (3, 1, 1)^t$ and $\mathbf{a}_\bullet = (1, 2, 1)^t$, respectively (the first component is the threshold or independent term of the linear function), which assertion is **CORRECT**?

A) There are four decision regions because there are two weight vectors and it is a two-dimensional representation space
B) The weight vectors $\mathbf{a}_\circ = (2, -2, -2)^t$ and $\mathbf{a}_\bullet = (-2, 0, -2)^t$ define the same decision boundary than the weight vectors given in the question statement The decision boundary equation is: $\mathbf{a}_\circ^t\mathbf{y} = \mathbf{a}_\bullet^t\mathbf{y}$. In both cases, we have: $y_1 = 2$.
C) The weight vectors $\mathbf{a}_\circ = (1, 2, 1)^t$ y $\mathbf{a}_\bullet = (3, 1, 1)^t$ define an equivalent classifier to the one given in the statement Opposed decision regions.
D) The decision boundary is defined as a plane in $\mathbb{R}^3$ because the weight vectors are three-dimensional

27 $\boxed{\text{C}}$ (January, 2016) We have three different classifiers for a two-class problem in $\Re^2$. One classifier is formed by the linear functions: $g_1(y) = 2\,y_1 + y_2 + 3$ and $g_2(y) = y_1 + 2$. The second classifier is formed by: $g_1'(y) = -2\,y_1 + y_2 - 1$ and $g_2'(y) = -y_1 + 2\,y_2$. The third classifier is formed by: $g_1''(y) = -2\,y_1 - y_2 - 3$ and $g_2''(y) = -y_1 - 2$. Which assertion is TRUE?

A) (g_1, g_2) y (g_1', g_2') are equivalent, but (g_1, g_2) y (g_1'', g_2'') are not.
B) (g_1, g_2) y (g_1', g_2') are not equivalent, but (g_1, g_2) y (g_1'', g_2'') are equivalent.
C) (g_1, g_2) y (g_1', g_2') are not equivalent, but (g_1', g_2') y (g_1'', g_2'') are equivalent. Common boundary $y_2 = -y_1 - 1$ but $R \neq R' = R''$
D) The three classifiers are not equivalent to each other.

28 $\boxed{\text{C}}$ (January, 2016) The figure on the right shows two bi-dimensional samples in 2 classes: $(x_1,\circ)$ and $(x_2,\bullet)$. Given the weight vectors $\mathbf{a}_\circ = (0, 1, -2)^t$ and $\mathbf{a}_\bullet = (0, 0, 1)^t$, if we apply the Perceptron algorithm only to the sample x_1, we obtain the new weight vectors $\mathbf{a}_\circ = (1, 1, -2)^t$ and $\mathbf{a}_\bullet = (-1, 0, 1)^t$. Which is the value of the learning factor α and margin b?



A) $\alpha = 1.0$ y $b = 0.0$
B) $\alpha = -1.0$ y $b = 0.5$
C) $\alpha = 1.0$ y $b = 0.5$
D) It is not possible to determine the value of α and b

Problems

1. Let be a classification problem among 3 classes, $c = \{A, B, C\}$, where the objects are represented using a vectorial space of three dimensions. The classifier is based on Linear Discriminant Functions (LDF):

$$g_c(\mathbf{x}) = \mathbf{w}_c \cdot \mathbf{x} \qquad \text{for every class } c$$

where $\mathbf{w}_c$ y $\mathbf{x}$ are represented in compact notation; that is: $\mathbf{w} = (w_0, w_1, w_2, w_3)^t \in \mathbb{R}^4$ and $\mathbf{x} = (x_0, x_1, x_2, x_3)^t \in \mathbb{R}^4$, con $x_0 = 1$. Taking into account:

$$\mathbf{w}_A = (1, 1, 1, 1)^t \qquad \mathbf{w}_B = (-1, 0, -1, -2)^t \qquad \text{y} \qquad \mathbf{w}_C = (-2, 2, -1, 0)^t$$

Solve the following:

a) Classify the point $\mathbf{x}' = (2, 1, 2)^t$.
b) We know that the point $\mathbf{x}' = (-1, 0, -1)^t$ belongs to the class A. Which values will $\mathbf{w}_A$,$\mathbf{w}_B$ and $\mathbf{w}_C$ have after the application of the Perceptron algorithm for this particular point using a learning rate $\alpha = 0.1$?
c) Given the point $\mathbf{x}' = (1, -1, 2)^t$ that belongs to class C, obtain one of the possible values of the LDFs that will classify it correctly

Solution:

a) Discriminant functions for $(2,1,2)^t$:

$$\begin{aligned} g_A(\mathbf{x}) &= \mathbf{w}_A \cdot \mathbf{x} = 6 \\ g_B(\mathbf{x}) &= \mathbf{w}_B \cdot \mathbf{x} = -6 \\ g_C(\mathbf{x}) &= \mathbf{w}_C \cdot \mathbf{x} = 1 \end{aligned}$$

Classification:

$$c(\mathbf{x}) = \arg\max_c \ g_c(\mathbf{x}) = A$$

b) Discriminant functions for $(-1,0,-1)^t$:

$$\begin{aligned} g_A(\mathbf{x}) &= \mathbf{w}_A \cdot \mathbf{x} = -1 \\ g_B(\mathbf{x}) &= \mathbf{w}_B \cdot \mathbf{x} = 1 \\ g_C(\mathbf{x}) &= \mathbf{w}_C \cdot \mathbf{x} = -4 \end{aligned}$$

Since $\mathbf{x}$ belongs to class A, the Perceptron algorithm modifies the weight vectors of the discriminant functions that return a value higher than g_A as well as the weight vector of g_A itself.

$$\begin{aligned} &\mathbf{w}_A^* = \mathbf{w}_A + \alpha\mathbf{x} = (1.1, 0.9, 1, 0.9)^t \\ &\mathbf{w}_B^* = \mathbf{w}_B - \alpha\mathbf{x} = (-1.1, 0.1, -1, -1.9)^t \\ &\mathbf{w}_C \text{ IS NOT MODIFIED} \end{aligned}$$

c) For example:

$$\begin{aligned} \mathbf{w}_A &= (0,0,0,0)^t \\ \mathbf{w}_B &= (0,0,0,0)^t \\ \mathbf{w}_C &= (-2,2,-1,0)^t \end{aligned}$$

2. Let be a classification problem among 3 classes, $c = \{1,2,3\}$, for objects represented by means of two-dimensional feature vectors. Given 3 training samples: $\mathbf{x}_1 = (0,0)^t$ belongs to class $c_1 = 1$, $\mathbf{x}_2 = (0,1)^t$ belongs to $c_2 = 2$, and $\mathbf{x}_3 = (2,2)^t$ belongs to $c_3 = 3$. Find a linear classifier with minimum error using the Perceptron algorithm. Set the initial weights to zero for all the classes, learning factor $\alpha = 1$ and margin $b = 0.1$. Show all the steps of the iterative algorithm, the value of the weights for all the classes, until convergence to the minimum error. Remember to use compact notation for the weights.

Solution:

Weight are shown in compact notation: $\mathbf{w}_c = (w_{c0}, w_{c1}, w_{c2})^t$

```
Iteration 1

Sample 1 belongs to class 1
g_1(x_1)=0
g_2(x_1)=0
Error: w_2=  -1   0   0
g_3(x_1)=0
Error: w_3=  -1   0   0
Error: w_1=   1   0   0
Sample 2 belongs to class 2
g_2(x_2)=-1
g_1(x_2)=1
Error: w_1=   0   0  -1
g_3(x_2)=-1.000000
Error: w_3=  -2   0  -1
Error: w_2=   0   0   1
Sample 3 belongs to class 3
g_3(x_3)=-4
g_1(x_3)=-2
Error: w_1=  -1  -2  -3
g_2(x_3)=2
Error: w_2=  -1  -2  -1
Error: w_3=  -1   2   1

Iteration 2
```

```
Sample 1 belongs to class 1
g_1(x_1)=-1
g_2(x_1)=-1
Error: w_2=  -2  -2  -1
g_3(x_1)=-1
Error: w_3=  -2   2   1
Error: w_1=   0  -2  -3
Sample 2 belongs to class 2
g_2(x_2)=-3
g_1(x_2)=-3
Error: w_1=  -1  -2  -4
g_3(x_2)=-1
Error: w_3=  -3   2   0
Error: w_2=  -1  -2   0
Sample 3 belongs to class 3
g_3(x_3)=1
g_1(x_3)=-13
g_2(x_3)=-5

Iteration 3

Sample 1 belongs to class 1
g_1(x_1)=-1
g_2(x_1)=-1
Error: w_2=  -2  -2   0
g_3(x_1)=-3
Error: w_1=   0  -2  -4
Sample 2 belongs to class 2
g_2(x_2)=-2
g_1(x_2)=-4
g_3(x_2)=-3
Sample 3 belongs to class 3
g_3(x_3)=1
g_1(x_3)=-12
g_2(x_3)=-6

Iteration 4

Sample 1 belongs to class 1
g_1(x_1)=0
g_2(x_1)=-2
g_3(x_1)=-3
```

3. Let be a classification problem between 2 classes, $c = \{1, 2\}$, where objects are represented by means of a two-dimensional feature vector. We have 2 training samples: $\mathbf{x}_1 = (0, 0)^t$ belongs to class $c_1 = 1$, and $\mathbf{x}_2 = (1, 1)^t$ belongs to class $c_2 = 2$. Find a minimum-error linear classifier by applying the Perceptron algorithm with initial weight vectors set to zero for the two classes, learning rate $\alpha = 1$ and margin $b = 0.1$. Show all the steps of the iterative algorithm and the successive updates of the weight vectors of the two classes, until convergence to the minimum error.

Solution:

Weight vectors are shown in compact notion: $\mathbf{w}_c = (w_{c0}, w_{c1}, w_{c2})^t$

```
Iteration 1

Sample 1 belongs to class 1
g_1(x_1)=0
g_2(x_1)=0
Error: w_2=  -1   0   0
Error: w_1=   1   0   0

Sample 2 belongs to class 2
g_2(x_2)=-1
g_1(x_2)=1
Error: w_1=   0  -1  -1
```

```
Error: w_2=   0   1   1

Iteration 2

Sample 1 belongs to class 1
g_1(x_1)=0
g_2(x_1)=0
Error: w_2=  -1   1   1
Error: w_1=   1  -1  -1

Sample 2 belongs to class 2
g_2(x_2)=1
g_1(x_2)=-1

Iteration 3

Sample 1 belongs to class 1
g_1(x_1)=1
g_2(x_1)=-1
```

4. We have a classification problem of two classes, A, B, where objects are represented by a two-dimensional feature vector. We have two training samples:

$$\mathbf{y}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in A, \quad \mathbf{y}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \in B,$$

   a) Set the initial weight vectors to 0 and show a trace of the Perceptron algorithm with learning rate $\alpha = 1.0$ and margin $b = 0.1$. Show the successive updates of the weight vectors and their final values.

   b) Obtain the equation of the decision boundary between the two classes according to the solution returned by the Perceptron algorithm. Represent graphically the boundary and the two training samples. Does this solution return a satisfactory classification?

Solution:

1. The algorithm executes two iterations of the main loop, yielding the next sequence of weight vectors:

$$\begin{array}{rcccccccccc}
\mathbf{y}: & & \mathbf{y_1} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \in A & & \mathbf{y_2} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \in B & & \mathbf{y_1} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \in A & & \mathbf{y_2} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \in B & & \\
g_A: & & 0 & & 1+2+2=5 & & 0-1+2=1 & & 0-2+1=-1 & & \\
g_B: & & 0 & & -1-2-2=-5 & & 0+1-2=-1 & & 0+2-1=1 & & \\
\text{error}: & & \mathit{true} & & \mathit{true} & & \mathit{false} & & \mathit{false} & & \text{SOLUTION} \\
\mathbf{a}_A: & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & +\,\mathbf{y_1} \;= & \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} & -\,\mathbf{y_2} \;= & \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} & \longrightarrow & \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} & \longrightarrow & & \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \\
\mathbf{a}_B: & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & -\,\mathbf{y_1} \;= & \begin{pmatrix} -1 \\ -1 \\ -2 \end{pmatrix} & +\,\mathbf{y_2} \;= & \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} & \longrightarrow & \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} & \longrightarrow & & \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}
\end{array}$$

2. Boundary decision equation:

$$0 - y_1 + y_2 = 0 + y_1 - y_2 \quad \rightarrow \quad y_2 = y_1$$

The graphical representation is a straight line passing through the origin with slope 45^0. The learning points are equidistant to both sides of the boundary, which can be considered an entirely satisfactory solution.