Lab Class ML:VI

By 2014-01-22 solutions for the following exercises have to be submitted: 1, 5b+c+d, 6a+b, 7.

Exercise 1 : Perceptron Learning

(a) Design a single perceptron with two inputs x_A and x_B. This perceptron shall implement the boolean formula $A \wedge \neg B$ with a suitable function $y(x_A, x_B)$.

(b) Train the perceptron from (a) with two iterations of the batch gradient descent, with a learning rate η of 0.1 and the weights initialized with $w_0 = -0.5$ and $w_1 = w_2 = 0.5$. Use the following examples in the given order:

x_1	x_2	$c(\mathbf{x})$
0	0	0
0	1	0
1	0	1
1	1	0

(c) Design a two-layer perceptron which implements the boolean formula A *XOR* B.

Use the values 0 for *false* and 1 for *true*, and the threshold function $\varphi(x) = \max(sign(x), 0)$.

Exercise 2 : Perceptron Learning

Given two single perceptrons a and b each of which defined by the inequality $w_0 + w_1x_1 + w_2x_2 \geq 0$. Perceptron a has the weights $w_0 = 1, w_1 = 2, w_2 = 1$, perceptron b has the weights $w_0 = 0, w_1 = 2, w_2 = 1$. Is perceptron a *more general* than perceptron b?

Exercise 3 : Neural Networks

Neural networks correctly classify only the feature vectors of a linearly separable training set D.

- ☐ false
- ☐ true
- ☐ true, if D is very large
- ☐ depends on the number of output units

Exercise 4 : Neural Networks

Which of the statements are true?

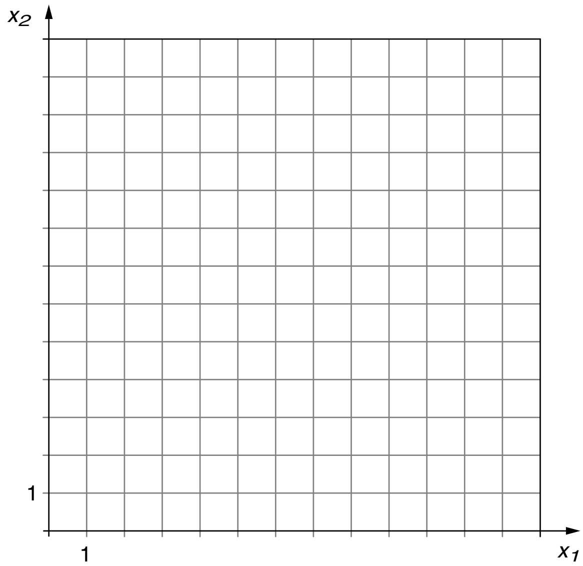
- ☐ Set-based gradient descent can approximate incremental gradient descent.
- ☐ Incremental gradient descent can approximate set-based gradient descent.
- ☐ Gradient descent finds always the global optimum.
- ☐ The perceptron algorithm converges always on linear separable data.

Exercise 5 : Perceptron Learning

Given is the following training data:

Number	1	2	3	4	5	6	7	8	9	10	11	12
x_1	1	7	7	2	3	2	5	10	11	12	10	9
x_2	3	4	6	6	2	1	7	3	5	8	5	8
$c(x_1, x_2)$	1	0	1	1	0	0	1	0	0	1	0	1

(a) Draw the data into the coordinate system bellow.



(b) The classes can be separated with straight lines. Determine such a straight line and specify its linear equation in the following form: $w_0 \cdot (-1) + w_1 \cdot x_1 + w_2 \cdot x_2 = 0$

(c) Given a perceptron that is initialized with the weights $w_0 = 1, w_1 = -1, w_2 = 3$. Which straight line belongs to this perceptron and how is it used for classification?

(d) Assuming the same initial weights, which weight changes are made by the incremental perceptron algorithm if at the learning rate $\eta = 0.1$, the whole dataset is presented in order three times? Compute the first three weight vectors by hand, and write a program in a language of your choice that solves all 36 iterations. Use the following table as a guide:

Iteration	x_1	x_2	$c(x_1, x_2)$	w_0	w_1	w_2	y	$Err(w)$
1	1	3	1	1	-1	3	?	?
2	7	4	0	?	?	?	?	?
3	7	6	1	?	?	?	?	?
⋮								
13	1	3	1	?	?	?	?	?
⋮								
36	9	8	1	?	?	?	?	?

Exercise 6 : Gradient Descent

(a) What are the differences between the perceptron training rule and the gradient descent method?

(b) What are the requirements for gradient descent being successful as a learning algorithm?

(c) What are the biggest problems of batch gradient descent in practice?

(d) What are the differences between the batch and the incremental (stochastic) gradient descent?

Exercise 7 : Perceptron Learning

Develop the gradient descend rule for a perceptron with output y, which applies the following special weight computation rule to its inputs:

$$y(x_1, \ldots, x_p) = w_0 + w_1(x_1 + x_1^2) + \ldots + w_p(x_p + x_p^2)$$

Like in the linear case, we look at the output of the first node of the perceptron without considering the threshold function.

Exercise 8 : Perceptron Learning

How can perceptrons be applied to solve a classification problem with more than two classes?