Aprendizagem 2023

Labs 6-7: Perceptron and Gradient Descent

Practical exercises

I. Perceptron

1. Considering the following linearly separable training data

	y_1	y_2	y_3	output
$\mathbf{x}_1$	0	0	0	–1
$\mathbf{x}_2$	0	2	1	+1
$\mathbf{x}_3$	1	1	1	+1
$\mathbf{x}_4$	1	-1	0	–1

Given the perceptron learning algorithm with a learning rate of 1 for simplicity, sign activation, and all weights initialized to one (including the bias).

a) Considering y1 and y2, apply the algorithm until convergence.
   Draw the separation hyperplane.
b) Considering all input variables, apply one epoch of the algorithm.
   Do weights change for an additional epoch?
c) Identify the perceptron output for $\mathbf{x}_{new} = [0\ 0\ 1]^T$
d) What happens if we replace the sign function by the step function?

$$\Theta\ (x)\ = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

   Specifically, how would you change the learning rate to get the same results?

2. Show graphically, instantiating the parameters, that a perceptron:

a) can learn the following logical functions: NOT, AND and OR
b) cannot learn the logical XOR function for two inputs

II. Gradient descent learning

Considering the following training data

	y_1	y_2	output
$\mathbf{x}_1$	1	1	1
$\mathbf{x}_2$	2	1	1
$\mathbf{x}_3$	1	3	0
$\mathbf{x}_4$	3	3	0

3. Let us consider the following activation

$$\hat{z} = output(\mathbf{x}, \mathbf{w}) = \frac{1}{1 + \exp(-2\mathbf{w} \cdot \mathbf{x})}$$

and half sum of squared errors as the loss function

$$E(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{N} (z_k - \hat{z}_k)^2 \quad where \;\; \hat{z}_k = output(\mathbf{x}_k, \mathbf{w})$$

a) Determine the gradient descent learning rule for this unit.

b) Compute the first gradient descent update assuming an initialization of all ones

c) Compute the first stochastic gradient descent update assuming an initialization of all ones.

4. Let us consider the following function:

$$output(\mathbf{x}, \mathbf{w}) = \frac{1}{1 + \exp(-\mathbf{w} \cdot \mathbf{x})}$$

and the cross-entropy loss function

$$E(\mathbf{w}) = -log(p(\mathbf{z}|\mathbf{w})) = -\sum_{k=1}^{N} (z_k log(\hat{z}_k) + (1 - z_k) log(1 - \hat{z}_k))$$

a) Determine the gradient descent learning rule for this unit

b) Compute the first gradient descent update assuming an initialization of all ones

c) Compute the first stochastic gradient descent update assuming an initialization of all ones

5. Let us consider the following function:

$$output(\mathbf{x}, \mathbf{w}) = \frac{1}{1 + \exp(-\mathbf{w} \cdot \mathbf{x})}$$

and half sum of squared errors as the loss function

a) Determine the gradient descent learning rule for this unit.

b) Compute the stochastic gradient descent update for input $\mathbf{x}_{new} = [1\ 1]^T$, $z_{new} = 0$ initialized with $\mathbf{w} = [0\ 1\ 0]^T$ and learning rate η=2

6. Consider the sum squared and cross-entropy losses. Any stands out?
What changes when one changes the loss function?