Neural Networks 2018
Homework 2
Perceptron learning rule vs. Delta rule

General instructions

Print your report and put it in the box of Neural Networks next to room 216 in Bernoulliborg. Also send a pdf version of the report by e-mail to neuralnets19+HW_2@gmail.com. The assignment is due on April 29^{th} at 11:00. You should e-mail just one archive file containing at least your report, so the easiest way of doing this is just to archive all of your files at once. Name it as follows:
s1234567_DonaldTrump_s7654321_SantaClaus.zip.
The maximum length of the report for this homework is **5 pages**.

Additional Information

For this assignment you can receive 8 points. If you receive p points, your homework grade is given by grade $= p \cdot \frac{10}{8}$, and it will be averaged with the lab assignment grade for this week.

1 Linear Algebra (2 pt.)

This section tests your linear algebra skills. The basic understanding of vectors, inner products and vector projections is essential to understanding neural networks. This assignment should not require more than 15 minutes. If it does require more, it is a good idea to (re)study some material on vectors. Take another look at section 3.2 of the book for example.

1. Draw a coordinate system on a piece of grid paper. Write 0 at the origin and depict the value +1 on both axes.
2. We consider two vectors [1]: $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{R}^2$

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \boldsymbol{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

3. Determine the length of $\boldsymbol{x}$ algebraically, denote this as $\|\boldsymbol{x}\|$ from hereon.
4. Determine the inner product of $\boldsymbol{x}^T\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y}$ algebraically.
5. Draw both vectors as arrows from the origin in your coordinate system. Use different colors or write the names of the vectors near the arrows. How can you check your answer for the previous question by looking at the geometry?
6. Let $\boldsymbol{w}$ be the weight vector. Draw $\boldsymbol{w}$ in the coordinate plane:

$$\boldsymbol{w} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

7. Draw a dotted line from the right end of $\boldsymbol{x}$ to the line $\boldsymbol{w}$ such that it is perpendicular to $\boldsymbol{w}$. The distance from the origin to the intersection of the dotted line and $\boldsymbol{w}$ is called the *projection* of $\boldsymbol{x}$ on $\boldsymbol{w}$. We write this as $\boldsymbol{x_w}$.

[1] Column vectors will be represented using bold font (LaTeX: `$\boldsymbol x$`) $\boldsymbol{x}$. Row vectors will be represented by adding an arrow at the top (LaTeX: `$\vec{ \boldsymbol x}$`): $\vec{\boldsymbol{x}}$.

8. Let ϕ be the angle between $\boldsymbol{x}$ and $\boldsymbol{w}$. Show that: $\cos(\phi) = \dfrac{\boldsymbol{x_w}}{\|\boldsymbol{x}\|}$.

9. Given the following relation: $\boldsymbol{x} \cdot \boldsymbol{w} = \|\boldsymbol{x}\|\|\boldsymbol{w}\|\cos(\phi)$, show that $\boldsymbol{x} \cdot \boldsymbol{w} = \boldsymbol{x_w}\|\boldsymbol{w}\|$.

10. Determine $\|\boldsymbol{w}\|$. Suppose that the threshold $\theta = 0.6$. Let $\boldsymbol{a}$ be an arbitrary vector. When does $\boldsymbol{a}$ satisfy $(\boldsymbol{a} \cdot \boldsymbol{w}) - \theta \geq 0$?

11. The plane is split in two by $\boldsymbol{w}$ and the threshold θ. Draw the separating line.

2 A TLU on paper (1.5 pt.)

1. Draw a simple artificial neuron with 2 inputs. Which elements can you distinguish other than the input and output? (Hint: don't only consider parameters, but also functions)

2. Which parts of the biologic neuron correspond to these elements?

3. Take a situation where both inputs to this artificial neuron are positive. How could hyperpolarization occur? Explain your answer.

3 Perceptron rule (1 pt.)

Take the TLU seen in Figure 1, which has the weight vector $\boldsymbol{w} = [0.6, 0.3]$ and threshold $\theta = 0.8$. The augmented weight vector is thus $\boldsymbol{w} = [0.6, 0.3, 0.8]$. The TLU is given an input vector $\boldsymbol{x} = [0.3, 1.0]$. The target t for this specific input is 1. Assume there is a learning rate of $\alpha = 0.25$. Show your working for the following calculations:

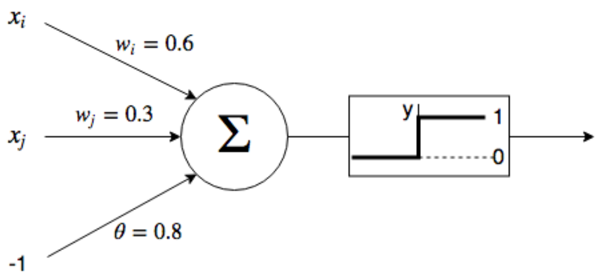

Figure 1: TLU with augmented weight vector $\boldsymbol{w} = [0.6, 0.3, 0.8]$.

1. Give the adjusted augmented weight vector $\boldsymbol{w}'$ for the given example.

2. Another input vector $\boldsymbol{x} = [0.6, 0.4]$ is fed in to the (newly updated) TLU. The target of this augmented input vector is $t = 0$. Use your newly adjusted weight vector to show how the perceptron rule would update it given input $\boldsymbol{x}$.

4 Delta rule (1 pt.)

When using a delta rule, the linear function $y = w_1 \cdot x_1 + w_2 \cdot x_2 + (-1) \cdot \theta$ is used as the output. We are thus using the activation of the TLU directly to calculate the error $e = (t - y) = (t - a)$ (Figure 2. This is unlike the perceptron rule where we used a thresholded output $y \in (0, 1)$.

Take the input vector $\boldsymbol{x} = [0.3, 1.0]$ and augmented weight vector $\boldsymbol{w} = [0.6, 0.3, 0.8]$. Just like in the previous exercise, the target t for this specific input is 1. Assume there is a learning rate of $\alpha = 0.25$. Show your working for the following calculations:

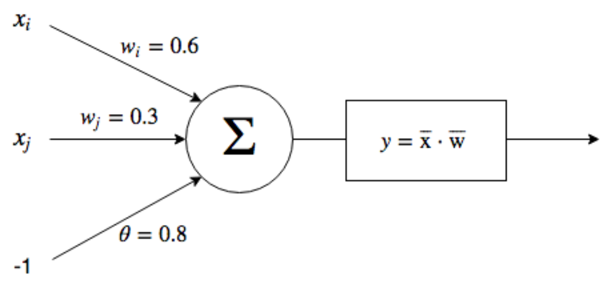

Figure 2: TLU with augmented weight vector $\boldsymbol{w} = [0.6, 0.3, 0.8]$.

1. Give the adjusted augmented weight vector $\boldsymbol{w}'$ for the given example.
2. Another input vector $\boldsymbol{x} = [0.6, 0.4]$ is fed in to the (newly updated) TLU. The target of this augmented input vector is $t = 0$. Use your newly adjusted weight vector to show how the delta rule would update it given input $\boldsymbol{x}$.

5 Logistic model (1 pt.)

A way to optimize the use of the delta rule is by using a logistic model. One such logistic model is $y = \sigma(w_1 \cdot x_1 + w_2 \cdot x_2)$, where σ is the sigmoid function such as the one seen in Figure 3, which constricts the output in the range $0 < y < 1$. The equation for such a function is:

$$y = \frac{1}{1 + e^{-(a-\theta)/\rho}}$$

Where a is the activation of the TLU, θ is the threshold of the TLU, and ρ is a constant used to set the steepness of the curve.

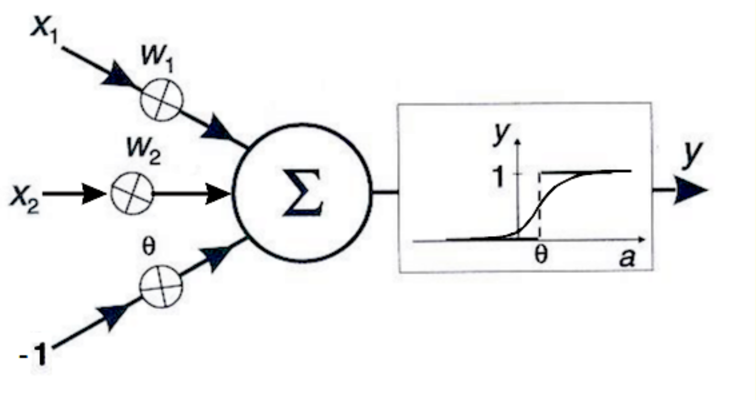

Figure 3: TLU with a sigmoid activation function. The activation of the neuron is fed into the sigmoid function, where the output of the function has a range $0 < y < 1$.

We will be using the same input vector $\boldsymbol{x} = [0.3, 1.0]$ and weight vector $\boldsymbol{w} = [0.6, 0.3]$ with a threshold $\theta = 0.8$. The target of this input vector is 1. Use a learning rate of $\alpha = 0.25$, and sigmoid function:

$$\sigma(a) = \frac{1}{1 + e^{-(a-\theta)/0.2}} \tag{1}$$

$$\sigma'(a) = \frac{d\sigma(a)}{da} = \frac{e^{(\theta-a)/0.2}}{0.2 \cdot (e^{(\theta-a)/0.2} + 1)^2} \tag{2}$$

Show your working for the following calculations:

1. Calculate $\boldsymbol{w}'$ and the new threshold for the given example.

2. Another augmented input vector $\boldsymbol{x} = [0.6, 0.4]$ is fed in to the (newly updated) TLU. The target of this augmented input vector is $t = 0$. Use your newly adjusted weight vector to show how the delta rule with a sigmoid activation function would update it given input $\boldsymbol{x}$.

6 Essay questions (1.5pt)

200 words maximum per question

1. What attributes make the mean-squared error a good error function?
2. What are the differences between the perceptron rule and delta rule?