LEIC-T 2023/2024
Aprendizagem - Machine Learning
Homework 3
Deadline 20/10/2024 20:00
Submit on Fenix as pdf

I) Polynomial Regression (3 pts)

Consider a training set with 5 observations (sample) with dimension $D = 1$

$$x_1=-0.8,\ x_2=1,\ x_3=-1.2,\ x_4=1.4,\ x_5=1.9$$

With targets

$$t_1=-20,\ t_2=20,\ t_3=-10,\ t_4=13,\ t_5=12$$

Consider as well the basis function

$$\phi_j(x) = x^j$$

which can lead to a polynomial regression of the third degree

$$y(x, \mathbf{w}) = \sum_{j=0}^{3} w_j \cdot \phi_j(x) = w_0 + w_1 \cdot x + w_2 \cdot x^2 + w_3 \cdot x^3.$$

(a) (1 pts)

Compute the design matrix Φ.

$$\Phi = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \phi_3(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \phi_3(\mathbf{x}_2) \\ \phi_0(\mathbf{x}_3) & \phi_1(\mathbf{x}_3) & \phi_2(\mathbf{x}_3) & \phi_3(\mathbf{x}_3) \\ \phi_0(\mathbf{x}_4) & \phi_1(\mathbf{x}_4) & \phi_2(\mathbf{x}_4) & \phi_3(\mathbf{x}_4) \\ \phi_0(\mathbf{x}_5) & \phi_1(\mathbf{x}_5) & \phi_2(\mathbf{x}_5) & \phi_3(\mathbf{x}_5) \end{pmatrix}$$

Φ =

$$\begin{pmatrix} 1. & -0.8 & 0.64 & -0.512 \\ 1. & 1. & 1. & 1. \\ 1. & -1.2 & 1.44 & -1.728 \\ 1. & 1.4 & 1.96 & 2.744 \\ 1. & 1.9 & 3.61 & 6.859 \end{pmatrix}$$

(b) (1 pts)

Compute the polynomial regression weights.

Solution:

By specifying as loss the sum of squared errors, we can minimize it by computing the gradient and check where it is zero. Doing so, yields the closed form solution for the weights:

$$\mathbf{w} = \left(\boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^T \mathbf{T}$$

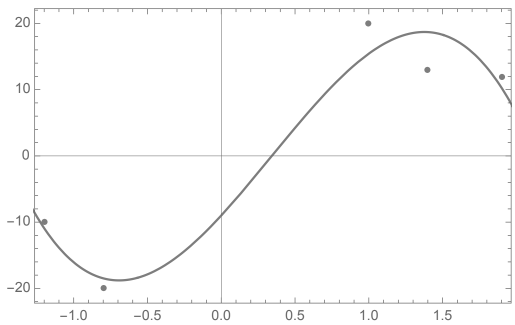
Applying it to the data we get:

$\Phi^T \Phi =$

$$\begin{pmatrix} 5. & 2.3 & 8.65 & 8.363 \\ 2.3 & 8.65 & 8.363 & 20.3569 \\ 8.65 & 8.363 & 20.3569 & 28.3232 \\ 8.363 & 20.3569 & 28.3232 & 58.8235 \end{pmatrix}$$

$w = (\Phi^T \Phi)^{-1} \Phi^T T = \{-8.9987, 24.219, 8.68547, -8.470627\}^T$

$y = -8.99873 + 24.2192 * x + 8.68547 * x^2 + -8.47063 * x^3$



(c) (1 pts)

LASSO regression (l1 regularization) lacks a closed form solution, why?

LEIC-T 2023/2024
Aprendizagem - Machine Learning
Homework 3
Deadline 20/10/2024 20:00
Submit on Fenix as pdf

Solution:

To apply the same process to this loss function we have to derive it. The leftmost term is the same as before. However, the rightmost term is non-differentiable at the origin since the l1 norm is the sum of absolute values of the weights and the absolute value's derivative is not defined at the origin.

II) Neural Network NN (4 pts)

Given the weights.:

$$W^{[1]} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$b^{[1]} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$W^{[2]} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$b^{[2]} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and the activation function ReLU

Rectifier also known as a ramp function

$$f(x) = \max(0, x). \tag{12.3}$$

is defined as the positive part of its argument [Jarrett *et al.* (2009)], [Nair and Hinton (2009)], [Goodfellow *et al.* (2016)]. The function is non-differentiable at zero; however, it is differentiable anywhere else and we can use the subderivative with sgn_0 function

$$f'(x) = sgn_0(x) = \begin{cases} 1 \text{ if } & x \geq 0 \\ 0 \text{ if } & x < 0 \end{cases}. \tag{12.4}$$

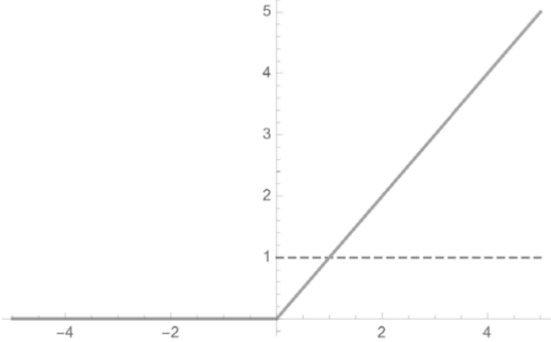
LEIC-T 2023/2024

Aprendizagem - Machine Learning

Homework 3

Deadline 20/10/2024 20:00

Submit on Fenix as pdf



of the hidden layer and SoftMax of the output layer using the cross entropy error loss do a stochastic gradient descent update (with learning rate η = 0.1) for the training example:

$$\mathbf{x}=(1,1,1,1,1)^T \text{ and the target } \mathbf{t}=(1,0)^T,$$

Solution

$$\mathbf{z}^{[1]} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix}, \quad \mathbf{x}^{[1]} = \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix}$$

$$\mathbf{z}^{[2]} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 15 \\ 15 \end{pmatrix}, \quad \mathbf{x}^{[2]} = softmax \begin{pmatrix} 15 \\ 15 \end{pmatrix} = \begin{pmatrix} \exp(15)/(\exp(15)+\exp(15)) \\ \exp(15)/(\exp(15)+\exp(15)) \end{pmatrix}$$

$$= \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

$$\delta^{[2]} = \frac{\partial E}{\partial \mathbf{x}^{[2]}} \circ \frac{\partial \mathbf{x}^{[2]}}{\partial \mathbf{z}^{[2]}} = (\mathbf{x}^{[2]} - \boldsymbol{t}) = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$\delta^{[1]} = \left(\frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{x}^{[1]}}\right)^T \cdot \delta^{[2]} \circ \frac{\partial \mathbf{x}^{[1]}}{\partial \mathbf{z}^{[1]}} = \mathbf{W}^{[2]^T} \cdot \delta^{[2]} \circ sgn_0(\mathbf{z}^{[1]}) = \mathbf{W}^{[2]^T} \cdot \delta^{[2]} \circ 1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\frac{\partial E}{\partial \mathbf{W}^{[1]}} = \delta^{[1]} \frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{W}^{[1]}} = \delta^{[1]} (\mathbf{x}^{[0]})^T = \begin{pmatrix} 0 \\ 0. \\ 0. \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

No change in weights of $W^{[1]}$ and in bias.

$$\frac{\partial E}{\partial \mathbf{W}^{[2]}} = \delta^{[2]} \frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{W}^{[2]}} = \delta^{[2]} (\mathbf{x}^{[1]})^T = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} ReLu(5,5,5) = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} (5,5,5) = \begin{pmatrix} -5/2 & -5/2 & -5/2 \\ 5/2 & 5/2 & 5/2 \end{pmatrix}$$

LEIC-T 2023/2024
Aprendizagem - Machine Learning
Homework 3
Deadline 20/10/2024 20:00
Submit on Fenix as pdf

$$\mathbf{W}^{[2]} = \mathbf{W}^{[2]} - \eta \frac{\partial E}{\partial \mathbf{W}^{[2]}} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} - 0.1 * \begin{pmatrix} -5/2 & -5/2 & -5/2 \\ 5/2 & 5/2 & 5/2 \end{pmatrix} = \begin{pmatrix} 1.25 & 1.25 & 1.25 \\ 0.75 & 0.75 & 0.75 \end{pmatrix}$$

$$\frac{\partial E}{\partial \mathbf{b}^{[2]}} = \delta^{[2]} \frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{b}^{[2]}} = \delta^{[2]} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$\mathbf{b}^{[2]} = \mathbf{b}^{[2]} - \eta \frac{\partial E}{\partial \mathbf{b}^{[2]}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0.1 * \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 0.05 \\ -0.05 \end{pmatrix}$$