Machine Learning for Bioinformatics

Exercise Sheet

Freie Universität Berlin, SoS 2024

Week 11 · Assignment on 12.06.2024. Submit until 21.06.2024 11pm.

Please note that the jupyter notebook must be submitted along with the exercise sheet![1]

Name:	Evita Vardhani	Matriculation no.:	5587165
Name:		Matriculation no.:	

Neural Network Architectures

Assignment 1. *Vanishing Gradient Problem*
The vanishing gradient problem can substantially impede the training of neural networks. There are several countermeasures that help to stabilize the gradient. This includes, for instance, activation functions that do not saturate at both very low and very high input values. Another strategy is to normalize the inputs of each layer, or to use skip connections.

1. Go through the first part of the notebook and complete the implementations. Points: 0

2. Which of the methods seems most efficient to obtain uniformly distributed magnitudes of the gradient, where at no point the gradient magnitudes drop to zero? Points: 2

Skip Connections:		Batch Norm:	✔
ReLU Activation:		ELU Activation:	

Assignment 2. *Enhancer Sequence Classification* Enhancers are regulatory elements in the genome that are responsible for cell type-specific gene expression. A major challenge is to understand in which cell type an enhancer becomes active. We would like study if the activity of an enhancer can be predicted from only the enhancer DNA sequence. It is believed that enhancers become active through transcription factors that bind to the DNA sequence. These proteins have binding subdomains that recognize specific short DNA subsequences within the enhancer. By constructing a classifier that predicts the activity from DNA sequence, we might be able to identify the DNA subsequences recognized by transcription factors. We restrict our attantion to enhancers that are active in liver (positive set) and try to discriminate them from those active in other tissues (negative set).

1.-3. Read and understand the Lightning implementations.

4. Complete the implementation of the CNN. Test different kernel sizes. Points: 0

[1] No points are awarded if an answer is only partially correct. Answers must be supported by results from the jupyter notebook.

5. What is the ROC-AUC of your classifier evaluated with 5-fold CV? You should get above 0.75. Points: 5

PR-AUC: 0.79

6. Visualize the kernels of the first layer of your CNN model to obtain the identified DNA subsequences (optional). Extra Points: 5

Assignment 3. *Pen&Paper Exercises*

1. The data

$$X = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

can be perfectly classified by a neural network with a single linear layer (including bias) and a sigmoid activation function to compute the output of the network. Points: 2

Correct: ☐ Incorrect: ✔

2. The output of a convolutional layer on images is rotation equivariant Points: 2

Correct: ✔ Incorrect: ☐

3. Gradient descent is an iterative algorithm for computing the minimizer x^* of an objective function f. At each iteration t it computes a value x_t based on the value x_{t-1} at the previous step $t-1$. Consider the objective function

$$f(x) = x^2 + 3\,.$$

Perform one iteration of the gradient descent algorithm with initial condition $x_0 = 1$ and step size $\gamma = 0.5$. What is the value of x_1? Points: 2

$x_1 =$ 1 - 0.5 (2x) = 1 - 0.5 (2.1) = 0

4. Assume we have a neural network with two layers and ReLU activation function. The first layer has weights

$$W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}$$

while the second layer uses the weight matrix

$$V = \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The output of the neural network is given by

$$f(X) = \sigma(\sigma(XW)V)$$

where X denotes the input and σ the ReLU activation function. Assume furthermore that our data is given by $X = \begin{bmatrix} 1 & 2 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \end{bmatrix}$.

Compute the value of the partial derivative of the loss function $L = (y - f(X))^2$ with respect to w_{11}. Points: 2

$\frac{\partial}{\partial w_{11}} L =$ 546

Maximum number of points: 15