Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2022/2023

Exercise Sheet 7

NNTI Tutorial

Deadline: 04.01.2023 08:00

Exercise 7.1 - Universal Approximation Theorem (0.5+0.5+0.5 points)

a) Please take a look at the Universal Approximation Theorem paper[1]. Please concisely cite the main contribution/idea of the paper.

b) Indeed there is a previous work[2] related to the representation capacity of neural networks. Why is this result weaker than the paper in Part a? What is the main idea of the paper?

c) State a reason and explain why, in practice, you would use deeper networks

Exercise 7.2 - L_1 and L_2 Regularization (0.25+0.5+0.5+0.25+0.5 points)

a) Explain why do we penalize only weights but not biases.

b) Both L_1 and L_2 regularization have their own advantages and disadvantages. In which case would you use L_1 regularization over L_2. Similarly, when would you prefer L_2 regularization over L_1? Explain your reasoning in detail.

c) Is it possible to use a new regularization method that makes use of both of L_1 and L_2? Motivate your answer and if yes what it would be beneficial for?

d) Regularization is defined as any modification that we make to a learning algorithm that is intended to reduce its generalization error as well as its training error. Is this true? Motivate your answer.

e) Why does L_1-regularization set variables to 0 and L_2 doesn't? Consider the case of L_2 and L_1-regularized least squares problem with 1 feature. Note that you should give mathematical justification.

Exercise 7.3 - Regularization Methods (0.5+0.5+0.5 points)

Consider the regularised objective $\tilde{J}(w) = J(w) + \frac{\lambda}{2} w^t w$

a) For SGD, the weight update for $J(w)$ is

$$w_{i+1} = w_i - \eta \nabla_w J(w)$$

[1] In case of a broken link: Horkin - Approximation Capabilities of Multilayer Feedforward Networks
[2] In case of a broken link: Cybenko - Approximation by Superpositions of a Sigmoidal Function

where η is the learning rate. Derive the weight update rule for the regularised loss. How is it related to weight decay?

b) Early stopping limits the number of updates that can be made to the parameters. Prove that it is equivalent to using L_2 regularisation.

c) Label smoothing is also a regularization method where we add noise ϵ to hard one-hot target like $[0, 1, 0, 0]$ to soften them. How would you modify a one-hot target vector with noise ϵ to perform label smoothing?

Exercise 7.4 - Dropout (0.5+0.5+0.5+0.5 points)

Dropout discussion. You are encouraged to research your answers by consulting the books or other online resources. Please keep your answers between 100 and 150 words.

a) What is dropout and how does it prevent overfitting? Mention a scenario where dropout is **not** too effective and what alternatives there are.

b) Explain how dropout is similar to and how it differs from traditional bagging.

c) Are other regularisation techniques—such as L_1 or L_2-regularization—mutually exclusive to dropout or can you apply both to your neural network? Explain why or why not. (You can use empirical results from research papers to support your answer).

d) Deep learning libraries typically implement a variant of dropout known as inverted dropout. How is standard dropout different from inverted dropout? What happens at inference time?

Exercise 7.5 - Regularization Implementation (1+2 points)

See attached notebook.

Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

- Please submit the assignment as a **team of two to three** students.
- Write the Microsoft Teams user name, student id and the name of each member of your team on your submission.
- Hand in zip file containing a `single` PDF with your solutions and the completed ipython notebook. Do not include any data or cache files (e.g. __pycache__).
- Important: please name the submitted zip folder and files inside using the format: **Name1_id1_Name2_id2**.
- Your assignment solution must be uploaded by only **one** of your team members to the 'Assignments' tab of the tutorial team (in **Microsoft Teams**). Please remember to press the **Hand In** button after uploading your work.
- If you have any trouble with the submission, contact your tutor **before** the deadline.