Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2023/2024

Exercise Sheet 9

Neural Networks

Deadline: 18.01.2024 23:59

Guidelines: You are expected to work in a group of 2-3 students. While submitting the assignments, please make sure to include the following information for all our teammates in your PDF/python script:

Name:

Student ID (matriculation number):

Email:

Your submissions should be zipped as **Name1_id1_Name2_id2_Name3_id3.zip** when you have multiple files. For assignments where you are submitting a single file, use the **same naming convention** without creating a zip. For any clarification, please reach out to us on the **CMS Forum**.
Note that the above instructions are mandatory. If you are not following them, tutors can decide not to correct your exercise.

Exercise 9.1 - Regularization (0.5 + 1 points)

a) Dropout is a regularization technique used in neural networks to prevent overfitting. Read the following paper which originally introduced the concept of Dropout and explain its key concepts as well as how it contributes to the prevention of overfitting during model training. [max 4 sentences].

b) Read the following paper and answer the following 3 question:

- What is *implicit regularization* and what is the role of the *implicit regularizer* in the context of Stochastic Gradient Descent (SGD)? [max 3 sentences]
- Discuss the modified loss function and explain how does the *implicit regularizer* affects the it when the learning rate is small and finite? [max 3 sentences]

Exercise 9.2 - Dropout (0.25+0.5+0.25+2.5 points)

a) Why is bagging typically not applied to neural networks and why can we use dropout instead? Also briefly explain the similarities between dropout and bagging.

b) Typically, frameworks implement a version of dropout known as *inverted dropout*. What is the relation between the two and what are the benefits of one over the other? Also discuss advantages at inference time (4-5 sentences).

c) Would you expect dropout to increase or decrease the capacity of a neural network? More precisely, if two networks have the same number of units, but one uses dropout and the other does not, which one would you expect to have higher capacity?

d) Now we explore the effect of dropout on regularization on a simple linear regression model trained using least squares. The basic linear regression model is given by

$$\hat{y}(\boldsymbol{x}, \boldsymbol{w}) = \sum_{i=1}^{m} w_i x_i$$

where the bias term is absorbed into the weight vector $\boldsymbol{w}$. With dropout, the sum-of-squares loss function becomes

$$J(\boldsymbol{w}; \{\boldsymbol{x}_i\}_{i=1}^{m}, \boldsymbol{y}) = \sum_{i=1}^{m} \left(y^{(i)} - \sum_{j=1}^{n} w_j R_{ij} x_j^{(i)} \right)^2 ,$$

where the elements $R_{ij} \in \{0, 1\}$ of the dropout matrix are chosen randomly from a Bernoulli distribution with parameter p. We now take an expectation over the distribution of random dropout parameters. Show that

$$\begin{aligned} \mathbb{E}[R_{ij}] &= p \\ \mathbb{E}[R_{ij} R_{ik}] &= \delta_{jk} p + (1 - \delta_{jk}) p^2. \end{aligned}$$

Using these results, show that the expected error for this dropout model is given by

$$\mathbb{E}_{\boldsymbol{R}}[J(\boldsymbol{w}; \{\boldsymbol{x}_i\}_{i=1}^{m}, \boldsymbol{y})] = \sum_{i=1}^{m} \left(y^{(i)} - p \sum_{j=1}^{n} w_j x_j^{(i)} \right)^2 + p(1-p) \sum_{j=1}^{n} w_j^2 \sum_{i=1}^{m} x_j^{(i)2}.$$

Congratulations, you have proved that the expected loss is a sum-of-squares loss with a quadratic regularizer in which the regularization coefficient is scaled separately for each input dimension according to the data values seen by that input!

Now, write down a closed-form solution for the weight matrix that minimizes this regularized error function.

Exercise 9.3 - Practical (5 points)

See the accompanying Jupyter Notebook.