Deep Learning

Theoretical Exercises – Week 6 – Chapter 6

Exercises on the book ”Deep Learning” written by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
Exercises and solutions by T. Méndez and G. Schuster

FS 2024

1 Exercises on Deep Feedforward Networks

1. Analyze the XOR problem:
   (a) Why is it not possible to solve the XOR problem with a single neuron?
   (b) To solve the XOR problem, a hidden layer is necessary. What property does the activation function of the hidden layer need for the network to be able to solve the XOR problem and why?
   (c) What is the task of the hidden neurons so that the network can solve the XOR problem?

   Solution:
   (a) A single neuron is a linear model and, thus, can only solve linearly separable classification problems. Since the XOR problem is not a linearly separable classification problem, a single neuron is not able to solve it.
   (b) The activation function of the hidden layer has to be nonlinear, since with a linear function the model would remain linear and thus could not solve the XOR problem.
   (c) The hidden neurons perform a nonlinear transformation of the input into a new feature space (representation), in which a linear model can solve the problem. Thus, the hidden layer performs the transformation and the output layer solves the now linearly separable classification problem.

2. What has to be considered when saturating activation functions are used in the output neurons or hidden neurons, respectively?

   Solution:
   A neuron with a saturating activation function causes the gradient to vanish and, thus, prevents the network from learning. For this reason, the loss should undo (compensate) the saturation in the output neurons.

Usually this means that the exponential function, which saturates when the argument is very negative, has to be compensated with the log function in the negative log-likelihood loss function.

In the hidden neurons, saturation can not be compensated. Hence, saturating activation functions should not be used in the hidden layers.

3. Assign the following keywords to the correct column. A keyword can be assigned to several columns.

- Gaussian output distribution
- Output of a probability
- Real-valued output
- Regression
- Output of a probability distribution
- Classification of four classes
- Multinoulli output distribution
- Minimizing mean-squared error
- Classification of two classes
- Maximizing log-likelihood
- Bernoulli output distribution
- Minimizing cross entropy

Solution:

Output Units		
Linear Units	**Sigmoid Units**	**Softmax Units**
• Gaussian output distribution • Real-valued output • Regression • Minimizing mean-squared error • Maximizing log-likelihood • Minimizing cross entropy	• Bernoulli output distribution • Output of a probability • Classification of two classes • Maximizing log-likelihood • Minimizing cross entropy	• Multinoulli output distribution • Output of a probability distribution • Classification of four classes • Maximizing log-likelihood • Minimizing cross entropy

Caution:
Minimizing the cross entropy always corresponds to maximizing the log-likelihood. However, Minimizing the mean-squared error only corresponds to maximizing the log-likelihood, if the output of the network is Gaussian distributed. The output is assumed to be Gaussian distributed, especially in linear regression.