LabMLISP: Lab Course Machine Learning in Signal Processing

Exercise III: Theory of neural networks

Prof. Veniamin Morgenshtern

Author: Matthias Sonntag

In this exercise, you are required to study the theory of neural networks and solve the exercises given below. We recommend the online book 'Neural Networks and Deep Learning' by Michael Nielsen. Pay special attention to Chapter 6 - Deep Learning, which deals with convolutional neural networks.

After completing this exercise, upload your solutions to StudOn. The solutions must be submitted as a PDF file. You may write the solutions using LaTeX and generate the PDF file, or you may write it on paper and upload a scanned copy of your solution.

Exercise 3.1: Perceptrons [Source: Chapter 1]

(a) Suppose we take all the weights and biases in a network of perceptrons, and multiply them by a positive constant, $c > 0$. Show that the behavior of the network doesn't change.

(b) Suppose we have the same setup as the last problem - a network of perceptrons. Suppose also that the overall input to the network of perceptrons has been chosen. We won't need the actual input value, we just need the input to have been fixed. Suppose the weights and biases are such that $wx + b \neq 0$ for the input x to any particular perceptron in the network. Now replace all the perceptrons in the network by sigmoid neurons, and multiply the weights and biases by a positive constant $c > 0$. Show that in the limit as $c \to \infty$ the behavior of this network of sigmoid neurons is exactly the same as the network of perceptrons. How can this fail when $wx + b = 0$ for one of the perceptrons?

Exercise 3.2: Convolutional neural networks (CNN)

(a) Draw the block diagram of a CNN that has characteristics given below. Label each layer of the CNN in the diagram and indicate the size of the corresponding tensor at the output of each layer. You may assume that the sizes are perfectly divisible by parameters, when needed.

- acts on images of size $N \times N$;
- has a convolutional layer that uses a kernel of size 5×5, stride length 1 and has M feature maps;
- uses a max-pooling layer with a 2×2 pooling window that is shifted across the feature maps with a stride length of 2;
- has a fully connected layer of K neurons which receives the flattened max-pool output;
- has a final layer which is also fully connected and returns an output that distinguishes 10 different classes.

- there are no bias terms in any layer.

(b) Calculate the total number of parameters for each layer of the network. First express the quantities in a general form, then set $N = 28$, $M = 20$, $K = 100$.

(c) Briefly explain the term “shared weights” in the context of CNN. How are they used in a CNN to detect different features in an image? Why does it not matter for the feature detection where in an image the feature is located (translation invariance of CNN)?

Exercise 3.3: Activation functions

(a) Give the equation for the sigmoid activation function, $\sigma(x)$. Draw the function for $x \in [-4, 4]$.

(b) Explain the vanishing gradient problem while training deep neural networks using sigmoid activation function.

(c) Give the equation for the rectified linear unit (ReLU) function, $\mathrm{ReLU}(x)$. Draw the function for $x \in [-4, 4]$.

(d) Give the equation for the softmax activation function of the i-th neuron, $\mathrm{softmax}(x)_i$. Using this equation, explain why softmax is often used in the output layer of a neural network that is used as a classifier.

Exercise 3.4: Reducing Overfitting

During training, a network ideally learns features from the limited set of available training data. To be useful, the network must be able work well on new data i.e., data unseen by the network during training. This is called *generalization*. The situation in which the network performs well on training dataset, but performs poorly on new data is called *overfitting*, and this is highly undesirable.
Briefly answer the questions given below in the context of training a neural network for the classification task.

(a) Explain the difference between the loss and the classification error.

(b) The data used for training neural networks is usually divided into training dataset, validation dataset and test dataset. Explain how each dataset is utilized in the training process.

(c) An effective method to reduce overfitting is to use more training data. Explain the term *data augmentation* and give at least three examples of data augmentation for images.

(d) Explain intuitively why regularization reduces overfitting in neural networks.

(e) Explain the idea of *dropout* method and how does it help with the overfitting problem.