LabMLISP: Lab Course Machine Learning in Signal Processing

Part III: The theory of deep learning

Prof. Veniamin Morgenshtern

Author: Matthias Sonntag

You next goal is to study the theory of deep learning and do some practice exercises. Under the link `https://machinelearning.tf.fau.de/course_mlisp.html`, go to Lectures 20–22. These lecture notes are taken from the online book by Michael Nielsen `http://neuralnetworksanddeeplearning.com` Read the lecture notes or the corresponding chapters in the online book. Pay special attention to Lecture 22, which is on convolutional neural networks.

Do the following exercises and submit the solutions to your tutors:

Problem 1: Lecture notes to Lecture 20, problem 1-1

Problem 2: Lecture notes to Lecture 20, problem 1-2

Problem 3: Convolutional neural networks (CNN)

1. Draw the diagram of a CNN that has following characteristics:
   - acts on images of size $N \times N$;
   - has M feature maps;
   - uses pooling with a 2×2 pooling window that is shifted across the feature maps with a stride length of 2;
   - has a densely connected layer of K neurons after the pooling layer;
   - distinguishes 10 different classes;
   - uses a kernel of size 5×5 and stride length 1 for convolution.

   Name each element of the CNN diagram and indicate the size of the corresponding tensor. You may assume that the sizes are perfectly divisible by parameters, when needed.

2. How many input connections does each neuron in the first densely connected layer has? How many input connections does each neuron in the output layer has? Assume the output layer is also densely connected. First derive the quantities in a general form, then set $N = 28$, $M = 20$, $K = 100$.

3. 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)?

Problem 4: Activation functions

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

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

Problem 5: Reducing Overfitting

During training, a network ideally learns features from the limited set of available training data. To be useful, the features learned in this way should be general enough to work well with many datasets (unseen by the network at train time). This is called *generalization*. The situation in which the network achieves low loss on the training set but performs poorly on the new data is called *overfitting*; this is highly undesirable. In this exercise consider the network built for classification. Answer briefly.

1. Explain the difference between the loss and the classification error.

2. Explain why during training it's a good idea to regularly verify the loss of the classifier on a dataset that is independent of the training set (commonly called the test set). Does the loss evaluated on the test set always represent the true classification performance?

3. 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.

4. Regularization has empirically been shown to improve generalization of neural networks. In this method the cost function is extended by a term that penalizes large weights or enforces some other structural properties. Explain intuitively why regularization reduces overfitting.

5. Explain the idea of *dropout* method and draw a relation to the ReLU activation function.