Exercise Sheet #7

Due: 2019-12-05, 16:00

Fred Hamprecht
Manuel Haußmann

General Regulations.

- You should hand in your solutions in groups of at least two people (recommended are three).
- The theoretical exercises can be either handwritten notes (scanned), or typeset using LaTeX.
- Practical exercises should be implemented in python and submitted as jupyter notebooks (`.ipynb`). Always provide the (commented) code as well as the output, and *don't forget to explain/interpret the latter!*
- Submit all your files in a single `.zip` archive to `mlhd1920@gmail.com` using the following standardized format: The subject line should consist of the full names of all team members as well as the exercise, and the title of the zip archive the last names. I.e. assuming your group consists of Ada Lovelace, Geoffrey Hinton and Michael Jordan, this means

  Subject: [EX07] Michael Jordan, Geoffrey Hinton, Ada Lovelace
  Zip Archive: `ex07-jordan-hinton-lovelace.zip`

1 Multilayer Perceptron (10 pt)

Throughout this work you will consider an MLP.

i) Encoding logic in an MLP. Assuming your input to the MLP is a D dimensional binary vector, $\mathbf{x} \in \{0,1\}^2$. How would you encode a logical AND (the output is equal to 1 iff[1] $x_i == 1 \ \ \forall i$), OR (the output is equal to 1 iff $x_i = 1$ for at least one x_i), and XOR (the output is equal to 1 iff $x_i = 1$ for exactly one i).

ii) Splitting of the feature space. Assuming you have a layer of M perceptrons (i.e. M lines). Assuming some regularity (i.e. no two decision boundaries are parallel and no three decision boundaries cross in a single point), in how many regions will they split the plane?

iii) An Example Split the feature space in Figure 1 optimally, and give the binary output vectors you get for each of the resulting regions. How would the corresponding region splitting a classification tree would produce (draw it once assuming axes aligned splits, and once sjuassuming obligue splits).

2 Training of an MLP (12 pt)

So far we haven't discussed in detail how to train a multilayer perceptron (MLP). For that reason we will rely on http://playground.tensorflow.org to nevertheless gain some intuition on what happens throughout the training of a neural network and how the different pieces interact and work together. For each part submit a screenshot and a short discussion of what you observed.

i) Fitting a Neural Net. Consider the spiral data set and the first two features (X_1, X_2). Come up with an architecture that can learn to classify the pattern well. You are free to use any number layers/neurons[2]/activation functions/regularization/...

[1] if and only if
[2] A neuron here is what we in the lecture talked about as a perceptron.

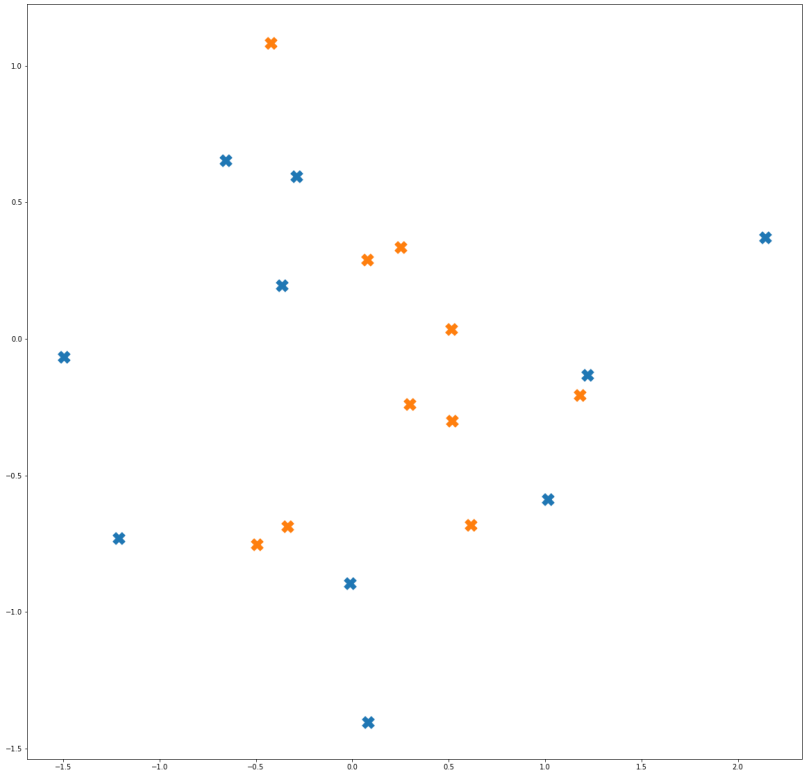

Figure 1: The feature space for the first exercise.

ii) Exploring Regularization. Pick the largest network size (i.e. 6 layers of 8 neurons each) and one of the data sets. Train it first without regularization, observing the behavior. Retrain it with L1 and L2 regularization and observe how the weight structure changes. What kind of behavior do you expect and does it fit with what you observe?

iii) Breaking Things. As discussed in the lecture, a net with enough parameters can fit any kind of pattern, even if there is none. Try to replicate this observation. Have your net learn a pattern "perfectly" (i.e. very low training error), but without having predictive power (i.e. the test error stays larger than random (which would be 0.5)). *Hint: If you consider the spiral data set with the minimal amount of training data and the maximum amount of noise, you get a collection of points with most structure removed.*