Exercises, May 3rd

Exercise 1

For the first exercise of the day, it is proposed that you come back to the demo software: https://playground.tensorflow.org, and we will use the spiral dataset again.

On Wednesday we ignored the regularization parameter. Today, we are going to evaluate the role this parameter can play. The input data – or features - will be the two coordinates X_1 and X_2. The final map will not be displayed in "Discrete Output Mode" in order to better understand how the output of the neural network varies.

We will keep the following hyper-parameters fixed:

- Number of hidden layers (3)
- Number of neurons per layer (8)
- There is a bias term in each neuron calculation.
- Learning rate: 0.03
- ReLU activation

1. How many neural networks parameters are fitted with these hyperparameters?

2. If we take the ratio of training to test equal to 20%, how many training data points do we have? Do you expect over-fitting?

3. Try one run to see what happens with the 20% ratio of training data if we calculate the neural network with no regularization. What do you observe?

4. Now we are going to run the program with different regularization parameters, and using L1 and L2 regularization. Write the mathematical expression of the loss function for the two configurations L1 and L2.

5. Recalculate the network for L2 optimization for values of the regularization parameter equal to 0.001, 0.003, 0.01 and 0.03. What do you observe in each case?

6. Now repeat the operation with seven instead of two input data (or features) : $X_1, X_2, X_1^2, X_2^2, X_1 X_2, \sin X_1, \sin X_2$, keeping all the hyperparameters unchanged. First calculate the network without regularization, then calculate it with a 0.03 regularization coefficient. What do you observe say after 1000 epochs?

Exercise 2

The article by (Rashka et al., 2018) discusses a number of approaches for optimizing the choice of hyperparameters. It is recommended that you read the whole paper but skip Paragraphs 1.7 (Confidence Intervals via Normal Approximations) and 2.4 (The Bootstrap Method and Empirical Confidence Intervals).

Questions about the text:

1. Please explain what is stratified sampling - or stratification - and why it is important.

2. Explain what is pessimistic and optimistic bias and how it relates to the size of the training and test set.

3. What are the three goals of performance estimation?

4. Why do we need a validation set?

5. What is the difference between holdout method, 2-fold cross-validation, and repeated holdout methods?