Deep Learning for Multimedia Pattern Recognition, Summer 2016 Exercise Sheet 2

Dr. Rolf Bardeli

April 25, 2016

1 Setting Up Deep Neural Networks

Write a function to set up a fully connected deep neural network. The function should take the following input parameters:

- Number of input neurons.
- Number of output neurons.
- List of hidden layer sizes (e.g. [1024, 256, 1024]).
- Activation function for the hidden neurons.
- Activation function for the output neurons.

2 A Toy Dataset

Set up the following extension of the XOR problem using random sampling.

- Create clusters of data points around the following centres:
  means = [(-2,2),(-1,2),(-1,1),(-2,1),(1,2),(2,2),(2,1),(1,1),
  (-2,-1),(-1,-1),(-1,-2),(-2,-2),(1,-1),(2,-1),(2,-2),(1,-2)]
- For each cluster centre, draw 400 times from the Gaussian distribution with the given centre and the respective covariance matrix from the following list:
  cov = [diag([0.1,0.1]), diag([0.15,0.07]), diag([0.15,0.07]), diag([0.1,0.1]),diag([0.1,0.1]), diag([0.15,0.07]), diag([0.15,0.07]), diag([0.1,0.1]),diag([0.1,0.1]), diag([0.15,0.07]), diag([0.15,0.07]), diag([0.1,0.1]),diag([0.1,0.1]), diag([0.15,0.07]), diag([0.15,0.07]), diag([0.1,0.1])]
- Create data labels by assuming that the n-th cluster has label n mod 2.
- Visualise the dataset.

3 Shallow vs. Deep

Train neural networks with different numbers of hidden layers and different sizes of hidden layers on the dataset produced in the previous exercise.

? What is the smallest size of a single hidden layer to produce a decent classification result?

? How does the training error depend on the number of neurons in a single hidden layer?

? How does the training error depend on the number of hidden layers?

? Are your results consistent with results from an independent test set? (Draw another dataset according to the same rules as in the last exercise.)

? Can you create a network with a very small total number of neurons but many hidden layers to achieve or exceed results with a large single hidden layer?

? Which activation function works best in which setting?

? What is better: a single output neuron or one for each class? Softmax or a different activation function in the output layer?