CST 463 - Advanced Machine Learning

Dr. Glenn Bruns

# Lab: Regularization

**Please work with a partner**. Try to answer without looking at the lecture nodes or other sources. If you can’t remember, discuss for a while with your partner before checking a source.

1. Why is learning rate scheduling needed with gradient descent? It already has a built-in adjustment based on the size of the derivatives. Try to come up with a couple of possible answers.
2. Geron writes "if you observe that the model is overfitting, you can increase the dropout rate". How do you tell if your neural net is overfitting?
3. Would it be possible to have the learning rate be a parameter in the network that is trained like any other parameter?
4. Dropout is like an ensemble method. Which ensemble method do you think it most closely resembles?
5. Geron writes "if you need a sparse model, you can add some regularization to the mix". What is a sparse model and why would you need it?
6. Is there a connection between learning rate and overfitting? For example, certain kinds of learning rate schedules be more likely to lead to overfitting?
7. We saw that some techniques that improve training have a computational cost, like leaky ReLU vs. ELU. What is the computational cost associated with dropout?
8. Try modifying your credit default model to include dropout.
9. If you still have time, answer this: minimization algorithms are trying to find a global minimum over a “mountain-like” surface that might have many local minima. Intuitively, how is regularization related to such a search?