Backpropagation in Practice

- **Use ReLU non-linearities**

- **Use cross-entropy loss for classification**

- **Use Stochastic Gradient Descent on minibatches**

- **Shuffle the training samples ($\leftarrow$ very important)**

- **Normalize the input variables (zero mean, unit variance)**

- **Schedule to decrease the learning rate**

- **Use "dropout" for regularization**