10.6 When to Use Deep Learning

Deep Learning vs Other Methods

- It seems like deep learning works really well.
  - . . . so it might be tempting to use it for everything.
- Here, we revisit the `Hitters` data from previous chapters.

Hitters Data

- Goal: predict `Salary` of a baseball player in 1987 using performance statistics from 1986.
- There are 263 complete cases, with 19 variables.
- The data are split into training (176 obs) and test (87 obs) sets.

Hitters Data: Models

- Three methods used:
  - Linear model with 20 parameters.
  - The same linear model, but with lasso regularization.
    - Tuning parameter set using 10-fold CV on training data.
    - This model has 12 nonzero coefficients.
  - Neural network with one hidden layer of 64 ReLU units.
    - This model has 1,345 parameters.

Hitters Data: Results

Model	Parameters	Mean Abs Error	Test Set R^2
Linear regression	20	254.7	0.56
Lasso	12	252.3	0.51
Neural Network	1345	257.4	0.54

- Similar performance for all three

Hitters Data

- Getting the configuration parameters set for the neural network is nontrivial.
- Running the linear regression and lasso models is easy.
- Linear regression and lasso also result in easy to present and interpret models.
  - Neural networks are essentially a black box.

When to Use Deep Learning

- Often the big successes occur when the signal to noise ratio is high, datasets are large, and overfitting is not a huge problem.
  - image recognition, language translation, ...
- For noisier data, simpler models often work better.
  - On the `NYSE` data, the AR(5) model is much simpler than a RNN, and performed as well.
  - In the `IMDB` data, the linear model fit by `glmnet` did as well as the neural network, and better than the RNN.
- In general, when faced with several methods that give roughly equivalent performance, we pick the simplest (Occam's razor)