BT5152 Tutorial 3

AY 2018/19, Semester 1, Week 5

Lu Wei

Flash Quiz

- Do you need to apply one-hot encoding on a column with values 0s and 1s?
- What’s the difference between a model’s parameters vs hyperparameters?
- If a model has accuracy 0.85 on training set and 0.79 on test set, is it overfitted?
- What’s a downside of min-max scaling?

Key Concepts Revision

Activation Functions

- Why is Sigmoid function used for logistic regression?
- Why do we need non-linear activation functions in hidden layers?
- Sigmoid vs. tanh vs. ReLU
- Is it OK to use a linear activation function on the output layer?

Error Functions (aka Loss Functions)

- Mean Squared Error

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

- Mean Absolute Error

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$$

- Root Mean Squared Error

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}$$

Formulas taken from wikipedia

Number of Nodes & Layers

- What should you start with?
- Why is a single hidden layer with sufficient nodes good enough?
- Why do we want multiple layers?
- How does it affect training speed?

Convergence & Stop Conditions

- Why random starting weights?
- How does learning rate affect training speed?
- How does learning rate affect model quality?
- On what scale should we be tuning the learning rate?
- Stop after x iterations (stepmax*) vs. stop at error threshold (threshold*)

* hyperparameters defined in the neuralnet R package

Hyperparameter Tuning

- Order of Tuning
  1. Learning rate
  2. Number of hidden units
  3. Number of hidden layers
- Hyperparameter Value Sampling Techniques
  1. Prefer random over grid search
  2. Coarse to fine

Recommended by Andrew Ng in Coursera course:
Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization

Tutorial Exercises:

RStudio > Console:

```
# install.packages("swirl")
library(swirl)
# delete_progress('your name')
install_course_github('weilu', 'BT5152', multi=TRUE)
swirl()
```

1: Neural Network