Evaluation and Credibility

How much should we believe in what was learned?

Introduction

- How predictive is the model we learned?
- Error on the training data is *not* a good indicator of performance on future data
  - ***Q: Why?***
  - A: Because new data will probably not be **exactly** the same as the training data!
- Overfitting – fitting the training data too precisely - usually leads to poor results on new data

Classification:
Train, Validation, Test split

Data

Results Known

+
+
-
-
+

Training set

Model Builder

Evaluate

Model Builder

Predictions

+
-
+
-

Validation set

+
-
+
-

Final Test Set

Final Model

+
-
+
-

Final Evaluation