Exercises in Machine Learning
Error Analysis

Zdeněk Žabokrtský, Ondřej Bojar
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University, Prague

Tue Apr 2, 2013
Outline

- Motivation for principled analysis.
- Bias vs. Variance.
- Optimizer vs. Objective function issue.
  - a.k.a. Search error vs. Modelling error.
- Error analysis, Ablative analysis.

Slides based on:
- the Stanford ML Lecture 11 http://www.youtube.com/watch?v=sQ8T9b-uGVE
- http://scott.fortmann-roe.com/docs/BiasVariance.html
- All errors are Ondřej’s fault.
Motivation

Some ML does not perform sufficiently well. You can consider random improvements:

- Getting more training examples.
- Reduce the set of features.
- Enlarge the set of features.
- Use different features.
- Run the optimizer for some more iterations.
- Choose a different optimization algorithm.
- Use a different regularization term or constant value.
- Try another learning algorithm (SVM).

... some may be fixing problems you don’t have.
Principled Analysis

First figure out what’s going on.

- Overfitting vs. Underfitting?
- Search error vs. Modelling error?
- Complex system: Find the most problematic component.

Trivial but vital:

- Visualize the data. (Plot or view frequent patterns.)
- Start with simple things.
Bias vs. Variance

High Variance = Overfitting:
- the model has too many parameters.

High Bias = Underfitting:
- the model is too rigid.

Consider:
- What is the effect of each of those on training error?
- Will more training data help?
- Sketch the shape of learning curves for each of those:
  - for the test error.
  - for the training error.
Bias-Variance Trade-off

\[ Err(x) = E[(Y - \hat{f}(x))^2] \]

can be decomposed as

\[ Err(x) = (E\hat{f}(x) - f(x))^2 + E(\hat{f}(x) - E\hat{f}(x))^2 + \sigma_e^2 \]

\[ Err(x) = \text{Bias}^2 + \text{Variance} + \text{Noise} \]

- Bias: how much the average predicted value differs from the ideal value
- Variance: how much a particular prediction differs from the average prediction


Derivation: see slides by Cohen
Bias-Variance Trade-off

Picture from: http://scott.fortmann-roe.com/docs/BiasVariance.html
Search vs. Modelling Error

Search Error:
- the optimizer fails to find the best parameters
- . . . a problem with the optimizer.

Modelling Error:
- the best parameters do not lead to the best performance.
- . . . a problem with the objective function.

Consider:
- Will more iterations help?
- When can two learners help to diagnose the problem?
Complex Systems

Error Analysis:
- Compares the best possible vs. current accuracy.
- Provide more and more golden truth data as part of the input.
- Find the component where the jump in accuracy is the highest.

Ablative Analysis:
- Compares some baseline vs. current accuracy.
- Switch off more and more components.
- Find the component where the loss in accuracy is the highest.
Steps Today

Warm up:

- plot PAMAP-easy dataset
- plot the fertility dataset in various ways
- find most frequent translations of “find”
  - find some highly frequent features of the case “find=zdát se”

Your learner (NB/KNN) on PAMAP or “find” data:

- plot the learning curves
  - of both training and test set
- discuss:
  - what are the options to increase accuracy
  - why does your learner perform well (or poorly)