Debugging Learning Algorithms

COMP 135 Intro to Machine Learning
Liping Liu

How to diagnose an algorithm?

- There is one way of being right, but there are many different ways of being wrong
  - Assumptions not hold in the application
    - Not i.i.d.,
    - \( x \) is NOT informative about \( y \) (\( p(y|x) \) is like random)
    - \( p(y|x) \) is NOT “smooth” in the space of \( x \)
  - Overfitting/Underfitting
  - Implementation issues

Diagnose a learning algorithm 1

- Q: is the problem with generalization to the test data?
- A: compare training error and test error (with a sense of variance)
  - High training error: do not expect to do better on test set
  - Low training error but high test error: the trained model does not generalize

Diagnose a learning algorithm 2

- Q: do you have train/test mismatch?
  - Do training and test sets have the same distribution?
- A: Shuffle the training and test data together and re-split
  - Error reduced: mismatch problem
  - Error not reduced: generalization problem

Diagnose a learning algorithm 3

- Q: Is your learning algorithm implemented correctly?
- A1: Is the loss or neg. llh “small” on training set or test set
- A2: Compare to existing packages
- A3: Sanity check with toy problems whose answer is known.
- A4: Copy the label as a feature and see if the classifier can discover it

Bias-Variance Trade-off

- Identify two parts of the error
  - Bias: because there is not a good model
  - Variance: because the learning algorithm does not find a good model

\[
\text{error}(f) = \text{error}(\hat{f}) - \min_{\hat{f}} \text{error}(\hat{f}) + \min_{\hat{f}} \text{error}(\hat{f})
\]

\[
\text{error}(\hat{f}) = \underbrace{\text{error}(\hat{f})}_\text{Variance: estimation error} + \underbrace{\text{error}(\hat{f})}_\text{Bias: approximation error}
\]

Error here is test error, NOT training error.
Bias-Variance Decomposition

• Almost impossible to separate the two errors in practice
• But we know the trend
  - Larger model space → Smaller approximation error (bias), larger estimation error (var)
  - Smaller model space → Larger approximation error (bias), smaller estimation error (var).

Relation to under/overfitting

• Underfitting ⇔ large approximation error (bias)
• Overfitting ⇔ large estimation error (variance)

Hyper-parameter and model space

• Almost every model has hyper-parameters controlling the size of model space
  - The size of model space is also called model complexity
• Two conceptual extremes
  - Lookup table: memorize training data and classify everything else as 0
  - Trivial classifier: classify everything as 0

Hyper-parameter and model space

• KNN classifier
  - $K$
  - Larger $K$ => smaller model space

Hyper-parameter and model space

• Linear classifier
  - $\lambda$
  - Larger $\lambda$ value => smaller model space

• Tree classifiers
  - Depth
  - Larger depth value => larger model space