CS 584 – Machine Learning

Topic: Bias vs Variance

Mustafa Bilgic

http://www.cs.iit.edu/~mbilgic

https://twitter.com/bilgicm

LEARNING

- True target
  - True distribution P
  - True function c
- Given a dataset D_1
  - Assume $D_1 \sim P$ is a representative sample
- We learn a model h_1 from D_1
- Error
  - What is the expected error of h_1 with respect to c?
- Previous lecture (3-classifier-evaluation.pdf)
  - True error/sample error
  - Accuracy, precision, recall, F1, AUC, …
  - Train/test split, train-val-test split, cross validation
- Another important question
  - How different would h be if I got a different dataset? For example, given D_2 and what would be the error of h_2 with respect to c?

A DIFFERENT SAMPLE *D*'?

- $D_1 \sim P$ is a finite sample
- What would happen if we got a different sample $D_2 \sim P$?
- For e.g., assume you need to estimate $P(Heads) = \theta$ and you collect a sample of 50 data points
  - Your D could be: $D_1 = \{23H, 27T\}$, $D_2 = \{22H, 28T\}$, $D_3 = \{27H, 23T\}$, etc.
- You estimate θ_i from D_i using your favorite method (e.g., MLE, Bayesian with different priors)
- Assume the true (unknown) parameter is θ^*
- What is the expected error of your estimates compared to the true θ^*
  - $E_D[(\theta_i - \theta^*)^2] = ?$

Bias-Variance Trade-off

- Given the set D of datasets D_i, models h_i, and true function c
- Compute the expected error
  - $E_D[(h - c)^2]$
  - In English: what is the expected error of the models h_i with respect to c, where the expectation is taken over possible samples (datasets)
- Why expectation over datasets?
  - I got one particular dataset, which is a sample of the true distribution, but I could've gotten another dataset, which is also a sample of the true distribution

Bias-Variance Trade-off Derivation

- $E_D[(h-c)^2] = Variance(h) + Bias^2(h,c)$
- See OneNote for derivation

EXAMPLES

- Single variable case
  - MLE vs Bayesian
- Regression
  - Polynomial regression with various degrees
  - Regularization
  - Neural network structures

Bias vs Variance

- High bias
  - More assumptions
    - E.g., naïve Bayes, linear regression, large priors
  - Simpler models
    - E.g., linear models, small decision trees, small neural networks, high regularization, …
- High variance
  - Fewer assumptions
    - E.g., logistic regression versus naïve Bayes, MLE vs Bayesian, small priors
  - More complex models
    - E.g., polynomial regression versus linear regression, bigger neural nets versus smaller ones, less regularization
  - Small data