Machine Learning Topic 3

Error Estimation

Aimee Barciauskas

29 March 2016

Error Estimation

- Definition of **empirical error**: *slide 1*
  - Empirical error may be misleading when $g = g_n$ is a data-based classifier, e.g. a data-based classifier may demonstrate 0 empirical error which is not representative of the true error of the classifier.
- LLN for $R_n(g)$: *slide 1*
- CLT for empirical deviation from true risk: *slide 1*
- Proof of Chebyshev: *slide 2*
- Typical deviations from expected values: *slide 3*
  - Chernoff Bounds: *slide 4*
    - Prime example for Chernoff Bounds: *slide 5*
  - Hoeffding's Lemma: *slide 5*
  - Hoeffding's Inequality: *slide 5*
  - Bernstein's Inequality: *slide 6*
    - Use to pick classifier from class of size N: *slide 6*
  - Bound on the true risk of data-based classifier from the empirical risk: *slide 7*
  - Bound on the true risk of data-based classifier from the best in class: *slide 7*
  - Proof the empirical risk minimizier is PAC: *bottom of slide 7-8*
- Definition of the empirical risk minimizer g_n: *slide 7*
- Bound on the true risk of data-based classifier when empirical risk is 0 and best in class is 0: *slide 9*
- Description of other error estimator (i.e. when all training data is used): *slide 10*
  - Leave-one-out, k-cross validation: *slide 11*

Notes

- Bounding of sums of random variables minus their expected values is used for bounding $|R_n(g) - R(g)|$: $R_n(g)$ is binomial a sum of n bernoullis, and $R(g)$ is it's expected value
- CLT gives bounds of the order $\frac{1}{\sqrt{n}}$
- Chebyshev gives us that typical deviations are of the order $\frac{\sigma}{\sqrt{n}}$
- Chernoff gives us something an exponentially decreasing upper bound that is non-asymptotic
- Hoeffding / Bernstein gives us something that is non-asymptotic and distribution-free