Sample-Based Scoring Rules

Reading Questions for April 13, 2016

We ask you to submit comments on the following papers by midnight Tuesday, April 12:

- Nonparametric Scoring Rules. E. Zawadzki and S. Lahaie AAAI'15.
- (optional) A Market Framework for Eliciting Private Data. B. Waggoner, R. Frongillo, and A. Abernethy, NIPS'15.

Your comments should include both answers to the specific reading questions and generic response about the papers. You are welcome to include any questions you have about the papers in your comments. After submitting your own comments, you'll be able to see others' submitted comments. You can comment on others' submissions and answer raised questions on Canvas. Discussion on Canvas is strongly encouraged.

1 Reading Questions

In these reading questions we'll focus on a concrete example. Let $\Omega = \mathbb{R}$, and let $K(x, \cdot)$ be a Guassian distribution of mean x and variance 1. So $K(x, \omega) = \frac{1}{\sqrt{2\pi}} e^{-(\omega - x)^2/2}$.

1. Intuitively, given the three samples $x_1 = -5, x_2 = 7, x_3 = 15$, what does the kernel density estimate $p_{K,X}(\omega)$, Equation 4, look like? How does this differ from the "empirical distribution" of the samples (Equation 5)?
2. Now let's imagine there's a "true" distribution P that is uniform on $[-20, 20]$. What happens to the kernel density estimate as we draw many samples from P? Similarly, how would you describe the relationship between the "sample score" (Equation 8) and the "kernel score" (Equation 6)?

3. In what scenarios would you prefer to use a sample-based scoring rule as proposed in this paper (Equation 8), versus a traditional proper scoring rule (e.g. Equation 6), for predicting a real-valued random variable?

2 Generic Response

Respond to the papers following the guidelines in the course syllabus (under "Submit Comments and Presenting Papers").