Active Learning with Statistical Models

Cohn, David A., Zoubin Ghahramani, and Michael I. Jordan. "Active learning with statistical models." ***Journal of artificial intelligence research*** **4 (1996): 129-145.**

What Problem does this paper try to solve?

This paper talks how one may select examples in a statistically optimal manner for statistically-based machine learning architectures, especially focused on supervised learning. It extends existing statistical approach on neural networks and proposes two alternative statistically-based learning architectures such as *mixtures of Gaussians* and *locally weighted regression.* The paper also presents the empirical results of applying active learning to these two new architectures and demonstrates how these architectures are efficient and accurate compared to a neural network.

How does it solve the problem?

The paper provides efficient data selection in mixtures of Gaussians and locally weighted regression.

- Mixtures of Gaussians: It uses EM(Expectation Maximization) algorithm to estimate means and variances from the training data. It selects training examples to minimize variance.
- Locally weighted regression(LWR): LWR can be seen as a variant of the LOESS(Locally estimated scatterplot smoothing) model, which performs a linear regression on points in the data set, weighted by a Gaussian kernel centered at a certain point x. It illustrates how to set the smoothing parameter of the Gaussian kernel to minimize variance.

derives statistically optimal criteria to minimize the learner's variance, selecting examples that reduce model uncertainty. The work compares these methods to neural networks, highlighting that, while optimal data selection for neural networks is computationally expensive, the methods proposed here are both efficient and accurate.

List of novelties/contributions

- It introduces optimal data selection to two architectures: mixtures of Gaussians and locally weighted regression.
- It shows that strategies of minimizing variance reduce model uncertainty for both models.
- It provides mathematical solutions for computing variance and expected error.
- It demonstrates significant improvements in efficiency and accuracy.

Downsides of the work

- Mixtures of Gaussians model may require to evaluate an excessive number of points to get good coverage of the potential query space in high-dimensional spaces.
- The strategy of minimizing variance ignores the bias component, which may lead to significant errors.