



Contribution ID: 43

Type: **not specified**

Which query strategy should one choose in active learning of real-valued functions when the model class is possibly misspecified?

Tuesday, 17 September 2024 10:05 (20 minutes)

We discuss the problem of active learning in regression scenarios. In active learning, the goal is to provide criteria that the learning algorithm can employ to improve its performance by actively selecting data that are most informative.

Active learning is usually thought of as being a sequential process where the training set is augmented one data point at a time. Additionally, it is assumed that an experiment to gain a label y for an instance x is costly but computation is cheap.

However, in some application areas, e.g. in biotechnology, selecting queries in serial may be inefficient. Hence, we focus on batch-mode active learning that allows the learner to query instances in groups.

We restrict ourselves to a pool-based sampling scenario and investigate several query strategies, namely uncertainty sampling, committee-based approaches, and variance reduction, for actively selecting instantiations of the input variables x that should be labelled and incorporated into the training set, when the model class is possibly misspecified.

We compare all active selection strategies to the passive one that selects the next input points at random from the unlabelled examples using toy and real data sets and present the results of our numerical studies.

Type of presentation

Talk

Classification

Mainly methodology

Keywords

Active Learning; Regression; Misspecification of Models.

Primary author: SPANGL, Bernhard (University of Natural Resources and Life Sciences, Vienna)

Presenter: SPANGL, Bernhard (University of Natural Resources and Life Sciences, Vienna)

Session Classification: Regression

Track Classification: Machine Learning