Leverage based sampling for classification

Contribution

We validate the results of leverage based sampling for LS-regression, shown by Ma et al. [1]. We explore the possibility of using the leverage based sampling distribution from LS-regression on 2 class classification, and introduce a two new methods for classification

Motivation

The importance of sampling methods are initiated by very large datasets where it is not feasible to use all of the available data. This is illustrated by the rise in online access to online video data. These data contain many frames that are basically the same and therefore redundant.

What is leveraging

A sampling scheme based on covariances between datapoints. We define a probability distribution representing the "importance" of each datapoint and then sampling from that.. Shown by Ma et al. [1] to perform better than uniform on