Data Mining: Learning from Large Data Sets - Fall Semester 2015

{usamuel, adavison, kurmannn}@student.ethz.ch

October 28, 2015

Large Scale Image Classification

Briefly describe the steps used to produce the solution. Feel free to add plots or screenshots if you think it's necessary. The report should contain a maximum of 2 pages. Keep in mind that you only have to submit one report per group. Please indicate the contribution of each group member to the project.

Problem formulation. The goal of this project was to use a SVM to classify a set of images into whether they are images of people, or images of nature. The images were provided in the form of pre-computed feature vectors.

Approach.

Mapper. Performs the online svm using *Stochastic Gradient Descent*. (passes sum of w on to reducer)

Reducer. Creates a weighted average of the w passed by the mappers.

Learning lambda. Cross validation

Results.

Workload distribution.