Approximate SVM Training
How to train a Support Vector Machine with 1 Million examples?

Properties of the SVM learning algorithm
» Takes a set of data samples
» Summarizes the set with samples that are essential
» Non-essential samples are not used in the resulting classifier
» Effectively splits the set into ‘useful’ and ‘useless’ samples

» Slow when used to learn from high-dimensional image data
» Acceptable training times with thousands of samples
» Does not easily scale to very large sets of image samples

Approximate training schemes
» Take multiple fixed-size instances of the SVM learning algorithm
» Arrange them to process manageable chunks of training data
» Different strategies possible: ‘remove useless’, ‘gather useful’, ...
» We provide a theory to reason about expected results, e.g.:
\[ \text{accuracy}(\text{SVM}(\text{SVM}(A) + B)) < \text{accuracy}(\text{SVM}(A + B)) \]
» Resulting classifiers are approximate - not an exact solution
» Allow for very large data sets, therefore overall better classifiers

Reduction scheme
» ‘Remove non-essential samples’ strategy
» Motivated by the following:
\[ \text{accuracy}(\text{SVM}(\text{SVM}(A) + \text{SVM}(B))) < \text{accuracy}(\text{SVM}(A + B)) \]
» Expected to be worse than an exact solution
» Larger chunks will give better results, reasoned by:
\[ \text{accuracy}(\text{SVM}(A)) < \text{accuracy}(\text{SVM}(B)), \text{ if } |A| < |B| \]
» Managed to reduce 1M samples to 43K samples in 10 hours

Incremental scheme
» ‘Accumulate essential samples’ strategy
» Can be shown to be more accurate than reduction:
\[ \text{accuracy}(\text{SVM}(\text{SVM}(A) + \text{SVM}(B))) < \text{accuracy}(\text{SVM}(A + B)) \]
» Combines well with ‘bootstrapping’
» For a given SVM capacity limitation, will extract maximum samples
» Extracted 5000 support vectors out of 1M in 17 hours

Data. 2500 to 1M eyes with linear transformations
Application. Eye-finder in a face detection scenario
Accuracy. Approximate incremental augmented set classifier performing significantly better than original 2500 sample SVM