How to overcome SVM memory requirement

Asked 4 years, 4 months ago Active 4 years, 4 months ago Viewed 1k times

I am using the SVM function (LinearSVC) in scikit-learn. My dataset and number of features is quite large, but my PC RAM is insufficient, which causes swapping, slowing things down. Please suggest how I can deal with this (besides increasing RAM).

In short, without reducing the size of your data or increasing the RAM on your machine, you will not be able to use SVC here. As implemented in scikit-learn (via <u>libsvm</u> wrappers) the algorithm requires seeing all the data at once.

partial_fit() method. One example of an online algorithm that is very close to SVC is the Stochastic Gradient Descent Classifier, implemented in sklearn.linear_model.SGDClassifier. Through its partial_fit method, you can fit your data just a bit at a time, and not encounter the sort of memory issues that you might see in a one-batch algorithm like SVC. Here's an example:

One option for larger datasets is to move to a model that allows online fitting, via the