Random Label Noise Cannot Elude Active Cleaning with SVM

Label noise examples can degrade the performance of machine learning algorithms and hence it is useful to identify them and take corrective action. It is shown in the literature that the majority of the label noise examples created by a random process are selected as a support vector of a two class SVM classifier. In this talk we show the potential reason for the success of that method. We demonstrate this by identifying the scenarios (geometrically) under which the majority of the label noise examples will not get selected as support vectors of the SVM classifier. It is easy to see that it is unlikely to have these scenarios created by a random process.