As a first baseline approach we implemented the parallel sgd for a normal linear svm. We beat the easy baseline by a good enough margin. Then we tried to apply the inverse kernel trick with the random fourier features presented in the lecture. We used an rbf kernel first, but didnt get any good results, then we tried a cauchy kernel, meaning we sampled from a laplace distribution to get our fourier features as described in the paper. Although we did improve, it was not by much and were still below our initial score with the linear svm. At some point we got desperate and just sampled our fourier features from any distribution we could find in the numpy documentation, student-t performed admirably well. But that was just kind of playing around because we didn’t even know what kernel we were approximating with that. Bottom line was, that we couldn’t beat the linear svm with our kernel approximation scheme. Judging from the score that other groups got there should be a kernel with which the data is linearly separable, but we just didn’t find it. It would be nice to get an “official” solution approach after the deadline.