PCA *doesn't* help on classification task

Aodong Li

December 6, 2017

This document only serves as informal personal thoughts.

When we utilize PCA to reduce the dimension of the dataset, we rely on the motivation that the direction with higher variance is more informative and could capture the rough structure of the data *in most cases*. But PCA doesn't guarantee that the dimension-reduced dataset can be classified better. The main reason is that PCA is agnostic to classes, i.e., it treats data of different classes blind to the underlying distribution (PCA is one method of descriptive statistics!).

The following figure shows that if we project the dataset onto the first principle component, then the two clouds of data are messed up and no longer separable.