Dimensionality Reduction

December 14, 2016

Abstract

This document introduces some fundamental notions of Dimensionality Reduction.

0.1 Dimensionality Reduction (Unsupdervised)

The goal of dimensionality reduction is to systematically and intelligently reduce the number of features considered in a dataset. Stated differently, trim the fat off. Often times, in one's eagerness to collect enough data for machine learning to be effective, you might add irrelevant features to your dataset. Bad features have the effect of hindering the machine learning process, and make your data harder to understand. Dimensionality reduction attempts to trim your dataset down to the bare essentials needed for decision-making.

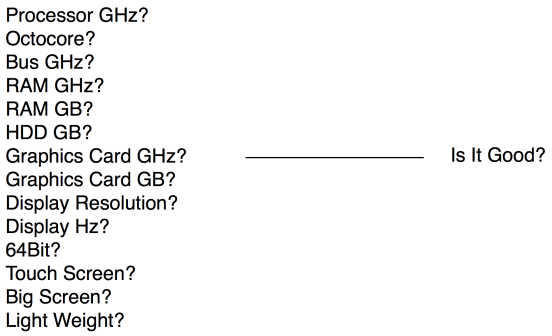

Figure 1: Example of dimensionality reduction.

More examples

- Given a 100 question survey, attempt to find the gist of what is truly being accessed; then rephrase it in just 5 questions.

- Build a robot that can recognize pictures of similar objects, even if they are rotated at odd angles and orientations.
- Compress a video stream by reducing the number of colors.
- Summarize a long book.

Dimensionality reduction falls into the realm of unsupervised learning because you don't instruct the computer which features you want it to build; the computer infers this information automatically by examining your unlabeled data.

source: course.edx.org