MITx 6.86x

Machine Learning with Python-From Linear Models to Deep Learning

7. Classification Using Manually Crafted Features

Bookmark this page

The performance of most learning algorithms depends heavily on the representation of
section, we will try representing each image using different features in place of the raw
Subsequently, we will investigate how well our regression model from the previous sect
different representations of the data.

Dimensionality Reduction via PCA

Principal Components Analysis (PCA) is the most popular method for linear dimension r
widely used in data analysis. For an in-depth exposition see: https://online.stat.psu.edu

Briefly, this method finds (orthogonal) directions of maximal variation in the data. By pr
dataset X onto $k \leq d$ of these directions, we get a new dataset of lower dimension th
variation in the original data than any other k-dimensional linear projection of X. By go
algebra, it can be proven that these directions are equal to the k eigenvectors correspo
eigenvalues of the covariance matrix $\widetilde{X}^T \widetilde{X}$, where $\widetilde{X}$ is a centered version of our ori

Remark: The best implementations of PCA actually use the Singular Value Decompositi
more straightforward approach outlined here, but these concepts are beyond the scop

Cubic Features

In this section, we will also work with a **cubic feature** mapping which maps an input ve
into a new feature vector $\phi(x)$, defined so that for any $x, x' \in \mathbb{R}^d$:

$$\phi(x)^T \phi(x') = (x^T x' + 1)^3$$

Discussion

Topic: Unit 2. Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Project 2: Digit recognition (Part 1) / 7. Classification Using Manually Crafted Features

posts

There are no posts in this topic yet.