PCA

WILLIAM MCCLOSKEY

Suppose we have an $m \times n$ design matrix X consisting of m mean-centered data points $x^{(1)}, \ldots, x^{(m)}$. PCA is a dimensionality-reduction technique which chooses a k-dimensional subspace $k < n$ and orthogonally projects the data onto that subspace. (Here k is a parameter chosen beforehand.)

PCA is usually formulated in two equivalent ways (e.g. in Bishop). The first is finding a subspace which maximizes the variance of the projected data. The second is finding a subspace which minimizes the reconstruction error of the original data. The straightforward equivalence between these two ideas seems not often made explicit, so we do that here.

For any subspace U, we can write the variance of the projected data as

$$\sum_{i=1}^{m} \|\mathrm{Proj}_U(x^{(i)})\|_2^2,$$

and we can write the reconstruction error as

$$\sum_{i=1}^{m} \|x^{(i)} - \mathrm{Proj}_U(x^{(i)})\|_2^2.$$

By orthogonality

$$\sum_{i=1}^{m} \|x^{(i)}\|_2^2 = \sum_{i=1}^{m} \|\mathrm{Proj}_U(x^{(i)})\|_2^2 + \sum_{i=1}^{m} \|x^{(i)} - \mathrm{Proj}_U(x^{(i)})\|_2^2,$$

so of course

$$\arg\max_U \sum_{i=1}^{m} \|\mathrm{Proj}_U(x^{(i)})\|_2^2 = \arg\min_U \sum_{i=1}^{m} \|x^{(i)} - \mathrm{Proj}_U(x^{(i)})\|_2^2.$$