Principal Components Analysis

Haoran Peng

June 2, 2021

Problem:

- Have a set of points $\{x^{(1)}, \ldots, x^{(m)}\}$ in $\mathbb{R}^n$
- Want to lossy compress them using function f, s.t. $f(x) = c \in \mathbb{R}^l$
- Then reconstruct them back using function g, so that $x \approx g(f(x))$. Let $g(c) = Dc$ where $D \in \mathbb{R}^{n \times l}$ have orthogonal columns

It is more often formulated as dimensionality reduction that maximizes variance of the data.

If we have one point, the objective is to minimize the L^2 norm:

$$\begin{aligned}
c^* &= \arg\min_c ||x - g(c)||_2 \\
&= \arg\min_c ||x - g(c)||_2^2 \\
&= \arg\min_c (x - g(c))^\top (x - g(c)) \\
&= \arg\min_c x^\top x - 2x^\top g(c) + g(c)^\top g(c) \\
c^* &= \arg\min_c -2x^\top g(c) + g(c)^\top g(c) \\
&= \arg\min_c -2x^\top Dc + c^\top D^\top Dc \\
&= \arg\min_c -2x^\top Dc + c^\top c
\end{aligned}$$

Solve by differentiation:

$$\begin{aligned}
\nabla_c - 2x^\top Dc + c^\top c &= 0 \\
-2D^\top x + 2c &= 0 \\
c &= D^\top x
\end{aligned}$$

This makes $f(x) = D^\top x$ and $g(f(x)) = DD^\top x = x$, a perfect reconstruction.

For all the points X (with stacked $x^{(i)}$), we minimize (where $D^T D = I$):

$$D^* = \arg\min_D ||X - XDD^\top||_F^2$$

Take the special $l = 1$ case (first principle component):

$$\begin{aligned}
d^* &= \arg\min_d ||X - Xdd^\top||_F^2 \\
&= \arg\min_d \mathrm{Tr}((X - Xdd^\top)^\top (X - Xdd^\top)) \\
&= \arg\min_d -\mathrm{Tr}(X^\top X dd^\top) - \mathrm{Tr}(dd^\top X^\top X) + \mathrm{Tr}(dd^\top X^\top X dd^\top) \\
&= \arg\min_d -2\mathrm{Tr}(X^\top X dd^\top) + \mathrm{Tr}(dd^\top X^\top X dd^\top) \\
&= \arg\min_d -2\mathrm{Tr}(X^\top X dd^\top) + \mathrm{Tr}(X^\top X dd^\top dd^\top) \\
&= \arg\min_d -2\mathrm{Tr}(X^\top X dd^\top) + \mathrm{Tr}(X^\top X dd^\top) \\
&= \arg\min_d -\mathrm{Tr}(X^\top X dd^\top) \\
&= \arg\max_d \mathrm{Tr}(X^\top X dd^\top) \\
&= \arg\max_d \mathrm{Tr}(d^\top X^\top X d) \\
&= \arg\max_d \frac{d^\top X^\top X d}{d^\top d}
\end{aligned}$$

This the the form of a Reyleigh quotient. The largest d is the eigenvector of $X^\top X$ that corresponds to the largest eigenvalue. Calculating the product $X \top X$ is sometimes not easy. We can instead do SVD on X:

$$\begin{aligned}
X &= U\Sigma V^\top \\
X^\top X &= V\Sigma^\top U^\top U\Sigma V^\top \\
X^\top X &= V\Sigma^2 V^\top
\end{aligned}$$

Which means V are the eigenvectors of $X^\top X$ and the projections are:

$$XV = U\Sigma V^\top V = U\Sigma$$

It is also standard to center data before doing PCA.