Circular coupled (sparse) PCA

Bharath Ananthasubramaniam

July 4, 2022

This approach aims to find two loading vectors (LVs) such that the projection of the data matrix on these two directions maximizes the projected variance and in addition principal components (PCs) satisfy an elliptical constraint.

Is this better formulated/called a type of SVD or a type of PCA.

We now formalize this idea. The data of interest is $\mathbf{X}$, a $n \times p$ matrix with p features and n samples. The coupled LVs are $\mathbf{v}_1$ and $\mathbf{v}_2$, each a $p \times 1$ vector, and the corresponding PCs are $\mathbf{u}_1$ and $\mathbf{u}_2$, each a $n \times 1$ vector. To obtain a variance interpretation for the projection, the data matrix $\mathbf{X}$ is column-centered, $\sum_j X_{ij} = 0$.

The general problem

We can formulate our approach as a reduced rank approximation similar to singular value decomposition, but with two key differences. First, we search for coupled PCs/SVs (two PCs simultaneously) that satisfy an elliptical constraint. Second, we do not enforce orthogonality of the two PCs.

$$
\begin{aligned}
&\min_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1} \|\mathbf{X} - \mathbf{u}_1\mathbf{v}_1^T - \mathbf{u}_2\mathbf{v}_2^T\|_F^2 = \\
&\quad = \min_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1} \operatorname{Tr}\left((\mathbf{X} - \mathbf{u}_1\mathbf{v}_1^T - \mathbf{u}_2\mathbf{v}_2^T)^T\ (\mathbf{X} - \mathbf{u}_1\mathbf{v}_1^T - \mathbf{u}_2\mathbf{v}_2^T)\right) \\
&\quad = \min_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1} \operatorname{Tr}\left[\mathbf{X}^T\mathbf{X} - \mathbf{X}^T\mathbf{u}_1\mathbf{v}_1^T - \mathbf{X}^T\mathbf{u}_2\mathbf{v}_2^T - \mathbf{v}_1\mathbf{u}_1^T\mathbf{X} + \mathbf{v}_1\mathbf{u}_1^T\mathbf{u}_1\mathbf{v}_1^T + \mathbf{v}_1\mathbf{u}_1^T\mathbf{u}_2\mathbf{v}_2^T\right. \\
&\qquad\qquad\qquad \left. -\mathbf{v}_2\mathbf{u}_2^T\mathbf{X} + \mathbf{v}_2\mathbf{u}_2^T\mathbf{u}_1\mathbf{v}_1^T + \mathbf{v}_2\mathbf{u}_2^T\mathbf{u}_2\mathbf{v}_2^T\right] \\
&\quad = \max_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1} 2\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + 2\mathbf{u}_2^T\mathbf{X}\mathbf{v}_2 - \|\mathbf{u}_1\|_2^2\|\mathbf{v}_1\|_2^2 - \|\mathbf{u}_2\|_2^2\|\mathbf{v}_2\|_2^2 - 2\mathbf{u}_1^T\mathbf{u}_2\mathbf{v}_1^T\mathbf{v}_2.
\end{aligned}
$$

We enforce a l_2-constraint on the LVs ($\|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1$) and the elliptical constraint ($\mathbf{u}_{1i}^2 + \mathbf{u}_{2i}^2 = 1\ \forall i \in \{1, \dots, n\}$).

$$
\max \left\{\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + \mathbf{u}_2^T\mathbf{X}\mathbf{v}_2 - \mathbf{u}_1^T\mathbf{u}_2\mathbf{v}_1^T\mathbf{v}_2\right\} \text{ s.t. } \|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1,\ \mathbf{u}_{1i}^2 + \mathbf{u}_{2i}^2 = 1\ \forall i \in \{1, \dots, n\} \tag{1}
$$

This transformation maps the data into the best possible approximate ellipse and the cost function maximizes the sum of the major and minor axis of this ellipse (note the sum is invariant to rotations of the ellipse).

Solving this optimization

This optimization is hard to solve; we need to either enforce orthogonality between the LVs or PCs, or optimize this coupled cost function. We simplify this problem by keeping only the first two terms in this optimization.

better way to justify this?

$$
\max \left\{\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + \mathbf{u}_2^T\mathbf{X}\mathbf{v}_2\right\} \text{ s.t. } \|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1,\ \mathbf{u}_{1i}^2 + \mathbf{u}_{2i}^2 = 1\ \forall i \in \{1, \dots, n\} (2)
$$

This simplified equation is bi-convex in $\mathbf{u}$s and $\mathbf{v}$s and can be solved using alternate maximization.

how do we prove this?

Given possible $\mathbf{v}_1$, $\mathbf{v}_2$, we can define dummy variables $\mathbf{y}_1 = \mathbf{X}\mathbf{v}_1$ and $\mathbf{y}_2 = \mathbf{X}\mathbf{v}_2$. The maximization for the $\mathbf{u}$s can be carried out for each i independently. We can maximize the cost function in (1) by choosing

$$\mathbf{u}_{1i} = \frac{\mathbf{y}_{1i}}{\sqrt{\mathbf{y}_{1i}^2 + \mathbf{y}_{2i}^2}} \quad \text{and} \quad \mathbf{u}_{2i} = \frac{\mathbf{y}_{2i}}{\sqrt{\mathbf{y}_{1i}^2 + \mathbf{y}_{2i}^2}} \tag{3}$$

Given now a solution from (3), the solutions for $\mathbf{v}$s that maximize the cost function can be computed independently simply as

$$\mathbf{v}_1 = \frac{\mathbf{X}^T\mathbf{u}_1}{\|\mathbf{X}^T\mathbf{u}_1\|_2} \quad \text{and} \quad \mathbf{v}_2 = \frac{\mathbf{X}^T\mathbf{u}_2}{\|\mathbf{X}^T\mathbf{u}_2\|_2} \tag{4}$$

We iterate (3) and (4) from a suitable random initial choice of vectors until convergence of $\mathbf{u}$s and $\mathbf{v}$s.

The sparse problem

We can obtain more interpretable solutions by constraining the LVs to be suitably sparse. This introduces a hyperparameter t controlling the sparsity of $\mathbf{v}$s.

$$\max \left\{\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + \mathbf{u}_2^T\mathbf{X}\mathbf{v}_2\right\}$$
$$\text{s.t. } \|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1,\ \|\mathbf{v}_1\|_1 \leq t, \|\mathbf{v}_2\|_1 \leq t,\ \mathbf{u}_{1i}^2 + \mathbf{u}_{2i}^2 = 1\ \forall i \in \{1, \ldots, n\} \tag{5}$$

Solving this optimization

This optimization (5) is solved in a similar manner to the general problem (1). Only the update equation (4) for the $\mathbf{v}$s differ as we also have to enforce the sparsity constraint. We solve the joint ℓ_1 and ℓ_2 constraint in a single step using the approach in Guillemot et al. (2019).