Principal Component analysis

Lukas Käll
lukask@kth.se

Decomposition of a matrix

- From linear algebra, we know that we can multiply two vectors into a matrix. $\mathbf{X} = \mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^{\mathsf{T}} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \\ u_4 v_1 & u_4 v_2 & u_4 v_3 \end{bmatrix}$.

Decomposition of a matrix

- From linear algebra, we know that we can multiply two vectors into a matrix. $\mathbf{X} = \mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^{\top} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \\ u_4 v_1 & u_4 v_2 & u_4 v_3 \end{bmatrix}$.
- What if we could do the opposite? I.e. given a matrix $\mathbf{X}$, what would be the vectors $\mathbf{u}$ and $\mathbf{v}$ that closest represents $\mathbf{X}$ so that $\mathbf{X} \approx \mathbf{u}\mathbf{v}^{\top}$. This is in essence, what you do with principal component analysis (PCA).

Conspicuous Example

- See notebook

More principal components to your PCA

- Once you remove the principal components from a matrix $\mathbf{X}$ the remaining residues,i.e. $\mathbf{X} - \mathbf{u}^{(1)}\mathbf{v}^{(1)\mathbf{T}}$ might in its turn be decomposed into vectors. I.e. we can calculate the vectors $\mathbf{u}^{(2)}$ and $\mathbf{v}^{(2)}$ that best describe the matrix $\mathbf{X} - \mathbf{u}^{(1)}\mathbf{v}^{(1)\mathbf{T}}$. We call these the second principal components, and the first are called first principal components.

More principal components to your PCA

- Once you remove the principal components from a matrix $\mathbf{X}$ the remaining residues,i.e. $\mathbf{X} - \mathbf{u}^{(1)}\mathbf{v}^{(1)\mathbf{T}}$ might in its turn be decomposed into vectors. I.e. we can calculate the vectors $\mathbf{u}^{(2)}$ and $\mathbf{v}^{(2)}$ that best describe the matrix $\mathbf{X} - \mathbf{u}^{(1)}\mathbf{v}^{(1)\mathbf{T}}$. We call these the second principal components, and the first are called first principal components.
- In this manner we can derive as many principal components as there are rows or columns (which ever is the lowest number) in $\mathbf{X}$. In most applications, we settle for two such components.

An illustration of PCA

		Sample					Eigensample
		1	2	3	$\dots$	M	$\mathbf{u}^{(1)}$
Gene	1	X_{11}	X_{12}	X_{13}	$\dots$	X_{1M}	$u_1^{(1)}$
	2	X_{21}	X_{22}	X_{23}	$\dots$	X_{2M}	$u_2^{(1)}$
	3	X_{31}	X_{32}	X_{33}	$\dots$	X_{3M}	$u_3^{(1)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$		$\vdots$
	N	X_{N1}	X_{N2}	X_{N3}		X_{NM}	$u_N^{(1)}$
Eigengene	$\mathbf{v}^{\mathbf{T}(1)}$	$v_1^{(1)}$	$v_2^{(1)}$	$v_3^{(1)}$	$\dots$	$v_M^{(1)}$	S_1

An illustration of PCA

		Sample					Eigensample	
		1	2	3	$\dots$	M	$\mathbf{u}^{(1)}$	$\mathbf{u}^{(2)}$
Gene	1	X_{11}	X_{12}	X_{13}	$\dots$	X_{1M}	$u_1^{(1)}$	$u_1^{(2)}$
	2	X_{21}	X_{22}	X_{23}	$\dots$	X_{2M}	$u_2^{(1)}$	$u_2^{(2)}$
	3	X_{31}	X_{32}	X_{33}	$\dots$	X_{3M}	$u_3^{(1)}$	$u_3^{(2)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$		$\vdots$	$\vdots$
	N	X_{N1}	X_{N2}	X_{N3}		X_{NM}	$u_N^{(1)}$	$u_N^{(2)}$
Eigengene	$\mathbf{v}^{\mathbf{T}(1)}$	$v_1^{(1)}$	$v_2^{(1)}$	$v_3^{(1)}$	$\dots$	$v_M^{(1)}$	S_1	
	$\mathbf{v}^{\mathbf{T}(2}$	$v_1^{(2)}$	$v_2^{(2)}$	$v_3^{(2)}$	$\dots$	$v_M^{(2)}$		S_2

Thanks!