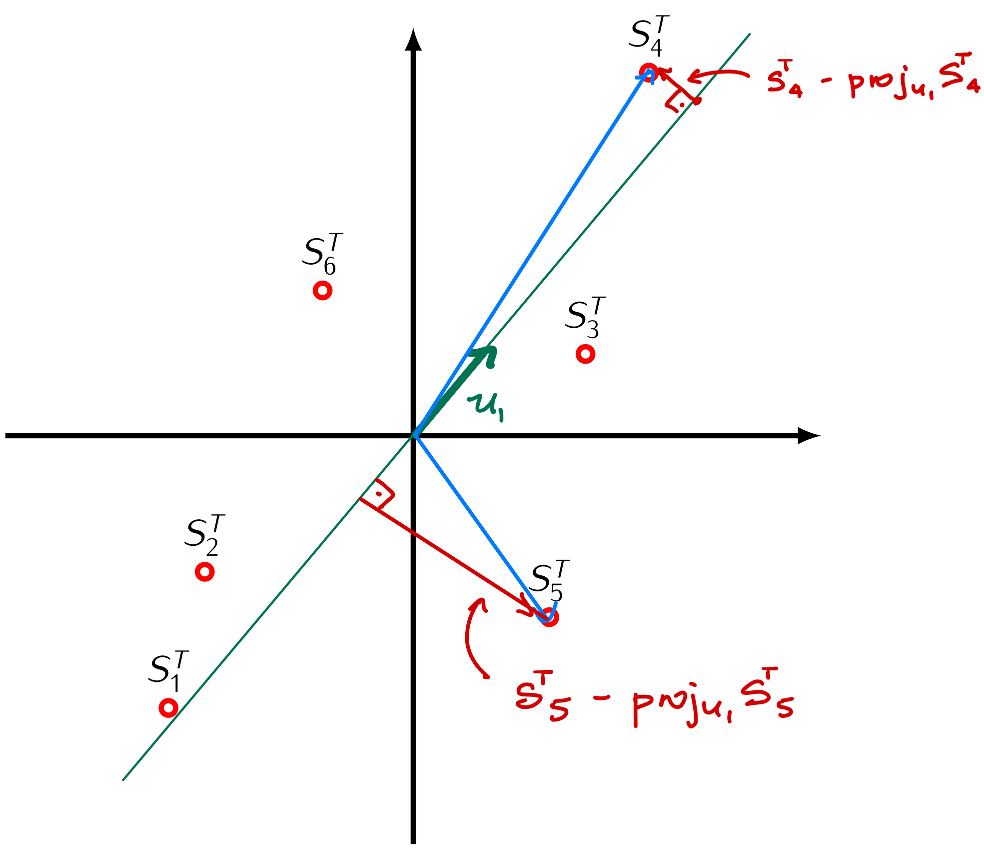
Example.

Demeaned data matrix:

$$A = \begin{array}{rc} & \begin{array}{cc}\text{Ex 1} & \text{Ex 2}\end{array} \\ \begin{array}{r}\text{Aly}\\ \text{Bob}\\ \text{Chen}\\ \text{Deb}\\ \text{Emma}\\ \text{Finn}\end{array} & \begin{bmatrix} -27 & -30 \\ -23 & -15 \\ 19 & 6 \\ 26 & 40 \\ 15 & -20 \\ -10 & 16 \end{bmatrix}\end{array} = \begin{bmatrix} \overbrace{X_1 \quad X_2}^{\text{feature vectors}} \end{bmatrix} = \left.\begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{bmatrix}\right\}\text{sample vectors}$$



Recall:

- The 1st principal axis of A is a vector $\mathbf{u}_1$ such that $||\mathbf{u}_1|| = 1$ and $\mathrm{Var}(A\mathbf{u}_1)$ is the largest possible.
- The vector $Y_1 = A\mathbf{u}_1$ is called the 1st principal component of A.

Note:

$$Y_1 = Au_1 = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{bmatrix} \cdot u_1 = \begin{bmatrix} s_1 u_1 \\ s_2 u_1 \\ \vdots \\ s_M u_1 \end{bmatrix} = \begin{bmatrix} s_1^T \bullet u_1 \\ s_2^T \bullet u_1 \\ \vdots \\ s_M^T \bullet u_1 \end{bmatrix}$$

dot product

Note: $\mathrm{proj}_{u_1} s_i^T = (s_i^T \cdot u_1)\, u_1$

$\|\mathrm{proj}_{u_1} s_i^T\| = |s_i^T \bullet u_1|$

$$\mathrm{Var}(Y_1) = \frac{1}{N}\sum_i (s_i^T \bullet u_1)^2 = \frac{1}{N}\sum_i |\mathrm{proj}_{u_1} s_i^T|^2$$

Upshot: u_1 is a unit vector such that the number $\sum_i |\mathrm{proj}_{u_1} s_i^T|^2$ is the largest possible.

Note: $|s_i^T|^2 = |\mathrm{proj}_{u_1} s_i^T|^2 + |s_i^T - \mathrm{proj}_{u_1} s_i^T|^2$

since $\mathrm{proj}_{u_1} s_i^T$ is orthogonal to $s_i^T - \mathrm{proj}_{u_1} s_i^T$

So: $\sum_i |\mathrm{proj}_{u_1} s_i^T|^2 = \sum_i |s_i^T|^2 + \sum_i |s_i^T - \mathrm{proj}_{u_1} s_i^T|^2$

We obtain: $\mathrm{Var}(Y_1) = \sum_i |\mathrm{proj}_{u_1} s_i^T|$ is the largest possible when $\sum_i |s_i^T - \mathrm{proj}_{u_1} s_i^T|$ is the smallest possible.

Proposition

Let A an $N \times M$ demeaned data matrix

$$A = \begin{bmatrix} \boxed{\quad S_1 \quad} \\ \boxed{\quad S_1 \quad} \\ \vdots \\ \boxed{\quad S_N \quad} \end{bmatrix}$$

• The 1^{st} principal axis of A is the vector $\mathbf{u}_1 \in \mathbb{R}^M$ such that $||\mathbf{u}_1|| = 1$ and the number

$$\sum_{i=1}^{N} ||S_i^T - \text{proj}_{\mathbf{u}_1} S_i^T||^2$$

is the smallest possible.

• The 1^{st} principal component of A is the vector

$$Y_1 = \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix}$$

such that $\text{proj}_{\mathbf{u}_1} S_i = c_i \mathbf{u}_1$ for $i = 1, \ldots, N$.