



## Why

We want to find a low-dimensional affine set into which we can project some high-dimensional data.

## Problem

For  $a \in \mathbf{R}^n$  and  $U \in \mathbf{R}^{n \times k}$ , the set  $W(a, U) = \{a + Uz \mid z \in \mathbf{R}^k\}$  is an affine set. Denote the projection of  $x \in \mathbf{R}^n$  onto  $W(a, U)$  by  $\text{proj}_{W(a, U)}(x)$ .

**Problem 1.** Given  $x^{(1)}, \dots, x^{(m)} \in \mathbf{R}^n$ , and a dimension  $k$ , find  $a \in \mathbf{R}^n$  and  $U \in \mathbf{R}^{n \times k}$  with  $U^\top U = I$  to minimize

$$\sum_{i=1}^m \|x^{(i)} - \text{proj}_{W(a, U)}(x^{(i)})\|^2,$$

the sum of squared distances between  $x^{(i)}$  and its projection on  $W(a, U)$ .

Express  $\text{proj}_{W(a, U)}(x)$  as  $UU^\top x + (I - UU^\top)a$  (see Projections on Affine Sets). We want to find  $a \in \mathbf{R}^n$  and  $U \in \mathbf{R}^{n \times k}$  to minimize

$$\sum_{i=1}^m \|x^{(i)} - UU^\top x^{(i)} - (I - UU^\top)a\|^2.$$

Fix  $U \in \mathbf{R}^{n \times k}$ . Define  $A \in \mathbf{R}^{nm \times n}$ ,  $B \in \mathbf{R}^{mn \times mn}$ , and  $\tilde{x} \in \mathbf{R}^{nm}$  by

$$A = \begin{bmatrix} I - UU^\top \\ \vdots \\ I - UU^\top \end{bmatrix}, \quad B = \begin{bmatrix} I - UU^\top & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & I - UU^\top \end{bmatrix}, \quad \text{and } \tilde{x} = \begin{bmatrix} x^{(1)} \\ \vdots \\ x^{(m)} \end{bmatrix}.$$

Then the objective is equivalent to

$$\|Aa - B\tilde{x}\|^2$$

Any minimizer  $a^*$  satisfies the normal equations

$$A^\top Aa^* = A^\top B\tilde{x}$$

Since  $(I - UU^\top)^\top = I - UU^\top$  and  $(I - UU^\top)^2 = I - UU^\top$ ,

$$A^\top A = \sum_{i=1}^m I - UU^\top = m(I - UU^\top)$$

and

$$A^\top B = \begin{bmatrix} I - UU^\top & \cdots & I - UU^\top \end{bmatrix}.$$

Consequently, we can express  $A^\top A a^\star = A^\top B \tilde{x}$  as

$$m(I - UU^\top) a^\star = \sum_{i=1}^m (I - UU^\top) x^{(i)}.$$

So  $a^\star$  is any vector satisfying

$$(I - UU^\top) a^\star = (I - UU^\top) (1/m) \sum_{i=1}^m (I - UU^\top) x^{(i)}.$$

One such point satisfying the above is  $\bar{x} = (1/m) \sum_{i=1}^m x^{(i)}$ . An expedient choice, as it does not depend on  $U$ .

Now we want to find  $U \in \mathbf{R}^{n \times k}$  to minimize

$$\sum_{i=1}^m \|(I - UU^\top)(x^{(i)} - \bar{x})\|^2.$$

Express the  $i$ th term of the sum as

$$\begin{aligned} \|(I - UU^\top)(x^{(i)} - \bar{x})\|^2 &= (x^{(i)} - \bar{x})(I - UU^\top)^\top (I - UU^\top)(x^{(i)} - \bar{x}) \\ &= (x^{(i)} - \bar{x})^\top (I - UU^\top)(x^{(i)} - \bar{x}) \\ &= \|x^{(i)} - \bar{x}\|^2 - \|U^\top (x^{(i)} - \bar{x})\|^2. \end{aligned}$$

The first term is a constant with respect to  $U$ . Define  $\bar{X} \in \mathbf{R}^{n \times m}$  by

$$\bar{X} = \begin{bmatrix} x^{(1)} - \bar{x} & \cdots & x^{(m)} - \bar{x} \end{bmatrix}.$$

Express the sum of the second terms by

$$\|U^\top \bar{X}\|_F^2 = \text{tr} \bar{X}^\top U U^\top \bar{X} = \text{tr}(U^\top \bar{X} \bar{X}^\top U).$$

So we seek  $U \in \mathbf{R}^{n \times k}$  with  $U^\top U = I$  to maximize

$$\text{tr}(U^\top \bar{X} \bar{X}^\top U).$$



