Elements of Machine Learning
Assigment 3 - Problem 2

Sangeet Sagar(7009050), Philipp Schuhmacher(7010127)
{sasa00001,phsc00003}@stud.uni-saarland.de

January 7, 2022

Problem 2 (T, 5+5 Points). PCA and PLS

1. Principal Components Analysis

The first principal component is the direction of maximum variance in the data. Show that this first principal component also minimizes the residual sum of squares, which is here the squared distance between the projected data point and the original data point.

Lets assume that $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n$ be n points with $x^{(i)} \in \mathbb{R}^d$.

Given n data points in d dimension

$$\boldsymbol{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{d \times n}$$

The goal is to reduce dimension from d to k. We choose k directions $u_1, u_2, \cdots, u_k$.

$$\boldsymbol{U} = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_k \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{d \times k}$$

For each u_j, we compute similarity $Z_j = \mathbf{u}_j^T \boldsymbol{X}$.

In vector form, we project $\boldsymbol{X}$ down to $\boldsymbol{Z} = (z_1, z_2, \cdots, z_k)^T = \boldsymbol{U}^T \boldsymbol{X}$. So first we encode i.e project $\boldsymbol{X}$ down to $\boldsymbol{Z}$ [Liang].

$$\begin{aligned} \boldsymbol{Z} &= (z_1, z_2, \cdots, z_k)^T \\ \boldsymbol{Z} &= (\mathbf{u}_1^T \mathbf{x}_1, \mathbf{u}_2^T \mathbf{x}_2, \ldots, \mathbf{u}_k^T \mathbf{x}_n) \\ \boldsymbol{Z} &= \boldsymbol{U}^T \boldsymbol{X} \end{aligned}$$

Then we project it back to determine the reconstruction error [Liang]

$$\begin{aligned} \tilde{\boldsymbol{X}} &= (\tilde{\mathbf{x}_1}, \tilde{\mathbf{x}_2}, \ldots, \tilde{\mathbf{x}_n}) \\ \tilde{\boldsymbol{X}} &= (\boldsymbol{U}\boldsymbol{U}^T \mathbf{x}_1, \boldsymbol{U}\boldsymbol{U}^T \mathbf{x}_2, \ldots, \boldsymbol{U}\boldsymbol{U}^T \mathbf{x}_n) \end{aligned}$$

Next we want the reconstruction error to be small

$$L(u) = \min_{\boldsymbol{U} \in \mathbb{R}^{d \times k}} \sum_{i=1}^{n} || \, \mathbf{x}_i - \boldsymbol{U}\boldsymbol{U}^T \mathbf{x}_i \, ||^2$$

Minimizing the reconstruction error

$$\mathcal{L}(U) = \sum_{i=1}^{n} \| x_i - UU^T x_i \|^2$$

$$= \sum_{i=1}^{n} (x_i - UU^T x_i)^T (x_i - UU^T x_i)$$

$$= \sum_{i=1}^{n} (x_i^T - x_i^T UU^T)(x_i - UU^T x_i)$$

$$= \sum_{i=1}^{n} (x_i^T x_i) - (\underbrace{x_i^T U U^T x_i}) - (\underbrace{x_i^T UU^T x_i}) + (x_i^T U U^T \underbrace{UU^T} x_i)$$

$$= \sum_{i=1}^{n} (x_i^T x_i) - 2(x_i^T UU^T x_i) + ((x_i^T U) U^T U (U^T x_i))$$

$$= \sum_{i=1}^{n} \underbrace{(x_i^T x_i)}_{\text{Scalar}} - 2 \underbrace{(x_i^T U U^T x_i)} + \underbrace{(U^T U)}_{\text{scalar qty : } U^T U = 1} (U^T x_i)^2$$

{use vector Property :- $A^T B = B^T A$ }

$= \cancel{\sum}$ ~~Product of two scalar qty~~

So we are left with :-

$$= \sum (U^T U)(U^T x_i)^2 - 2 (U^T x_i)^2$$

we need to focus on minimizing this term only :-

$$= -\sum_{i=1}^{n} (U^T x_i)^2 \quad ; \quad \|U\| = 1$$

or simply maximize $\sum_{i=1}^{n} (U^T x_i)^2$ = Same as variance of x_i when Projected on vector U.

$\Rightarrow \max U^T \Sigma U$ where $\Sigma = \left(\sum_{i=1}^{n} x_i x_i^T \right)$

hence mimizing the reconstruction error is same as maximizing the variance of projected points on vector U.

2. Partial Least Squares

Show that the first partial least squares direction solves

$$\begin{aligned} max_\alpha \quad & \mathrm{Cor}^2(y, \mathbf{X}\alpha)\mathrm{Var}(\mathbf{X}\alpha) \\ & \text{subject to } ||\alpha|| = 1 \end{aligned}$$

where Cor is Pearson's correlation coefficent. Thus, the PLS direction is a compromise between the least squares regression coefficient and the principal component directions.

References

Percy Liang. Lecture notes: Linear dimensionality reduction (practical machine learning (cs294-34)). `https://people.eecs.berkeley.edu/~jordan/courses/294-fall09/lectures/dimensionality/slides.pdf`. Published: 2009-09-24.