Recitation Week 9

Ashwin Bhola

CDS, NYU

Oct 30th , 2019

PCA

1. Suppose $x_1, \ldots, x_n \in \mathbb{R}^d$ are datapoints you want to represent in $k < d$ dimensions.
    1. Explain how to do this using PCA
    2. How can you implement PCA using SVD?
    3. How to determine an optimal value for k?

PCA

1. PCA steps:

   i. Center your data: If $\mathrm{X} = \begin{bmatrix} \cdots x_1 \cdots \\ \vdots \\ \cdots x_n \cdots \end{bmatrix} \in \mathbb{R}^{n \times d}$ is your data matrix, then $\mathrm{A}[i, \mathrm{j}] = X[i, j] - \frac{1}{n} \sum_p X\,[p, j]$

   ii. Use the new matrix A of centered data instances to construct the covariance matrix $S = A^T A \in \mathbb{R}^{d \times d}$

   iii. Spectral decomposition of S: $S = VDV^T$

   iv. Choose the top k eigenvalues $(\lambda_1, \ldots, \lambda_k)$ and the eigenvectors $(v_1, \ldots, v_k)$ from V corresponding to those eigenvalues

   v. Construct your new data matrix as $X_{new} = AV_k = \begin{bmatrix} \cdots a_1 \cdots \\ \vdots \\ \cdots a_n \cdots \end{bmatrix} \begin{bmatrix} \vdots & \vdots & \vdots \\ v_1 & \cdots & v_k \\ \vdots & \vdots & \vdots \end{bmatrix} \in \mathbb{R}^{n \times k}$

PCA

2. PCA using SVD: Let's do SVD on our centered data matrix A: $A = M\Sigma N^T$

$$\Rightarrow A^T A = N\Sigma^T \Sigma N^T = NDN^T$$

We see that $D = \Sigma^T \Sigma \Rightarrow \lambda_i = \sigma_i^2$

Thus we can use the first k right singular vectors to perform the projection

3. Optimal value of k can be chosen using
    1. Scree plot
    2. Fraction of variance explained by top k eigenvalues

PCA

1. Let $X \in \mathbb{R}^{n \times d}$ be your matrix of data points. Suppose you are implementing PCA. Someone suggests that you should standardize your data before calculating the eigenvalues. How do you standardize the data? Is it really required?

PCA

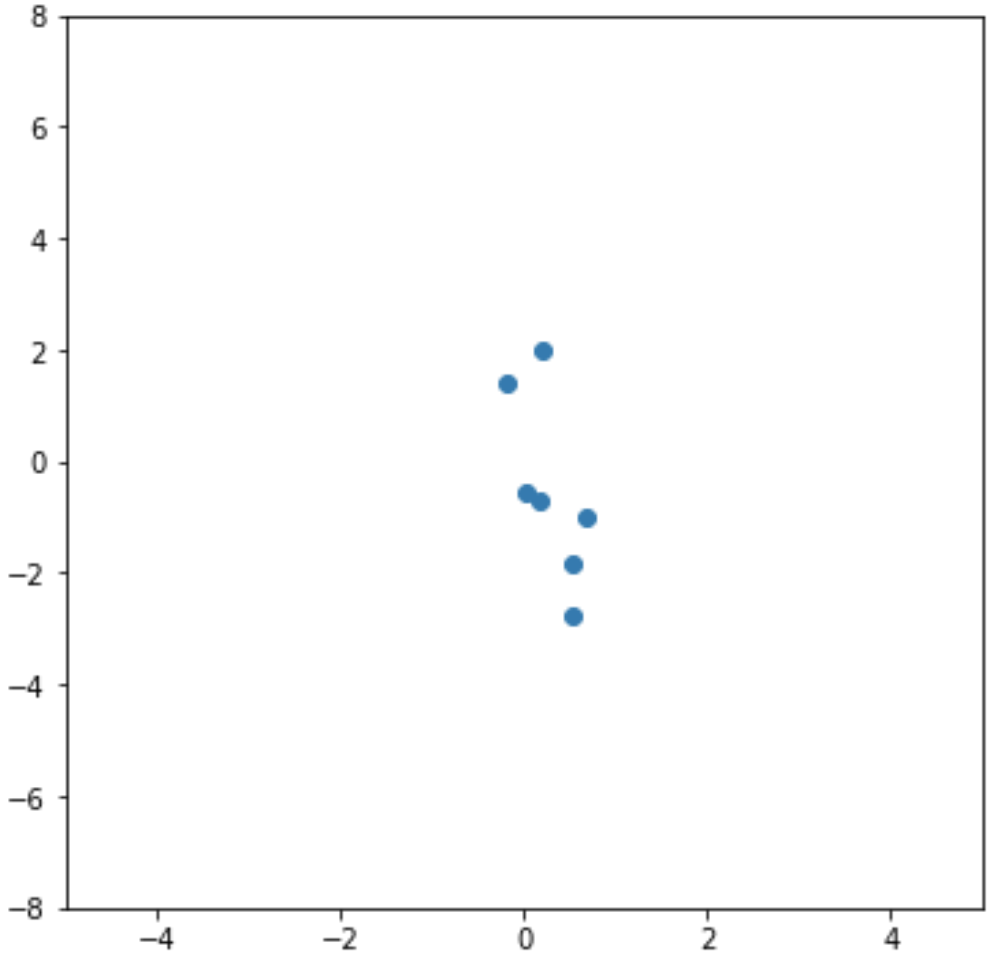
Solution: If $\mathrm{X} = \begin{bmatrix} \cdots x_1 \cdots \\ \vdots \\ \cdots x_n \cdots \end{bmatrix} \in \mathbb{R}^{n \times d}$ is your data matrix, then $\mathrm{A}[i,j] = \frac{X[i,j] - \mu_j}{\sigma_j}$, where

$$\mu_j = \frac{1}{n} \sum_p X\,[p,j]\ and$$

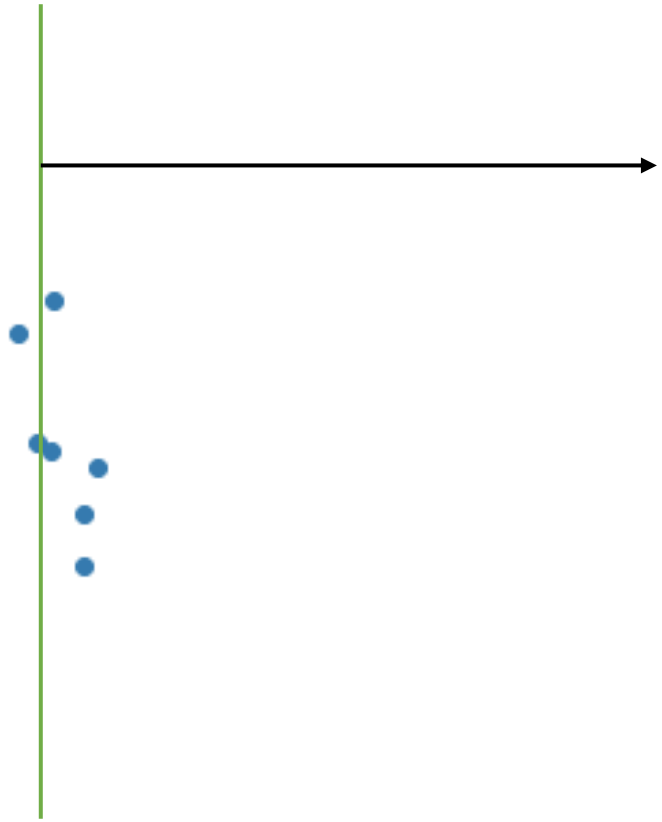
$$\sigma_j^2 = \sum_p \big(X[p,j] - \mu_j\big)^2$$

Centering the data instances is necessary because otherwise, the principal components get influenced by the location of the data in the feature space. This is equivalent to saying that if we don't center the data, then we are trying to approximate the data by a subspace and not an affine space which is less efficient. You may not want to standardize the data if you think the differences in scale are informative. In general, features with larger scale influences the principal components more which may not be desired in some scenarios

PCA

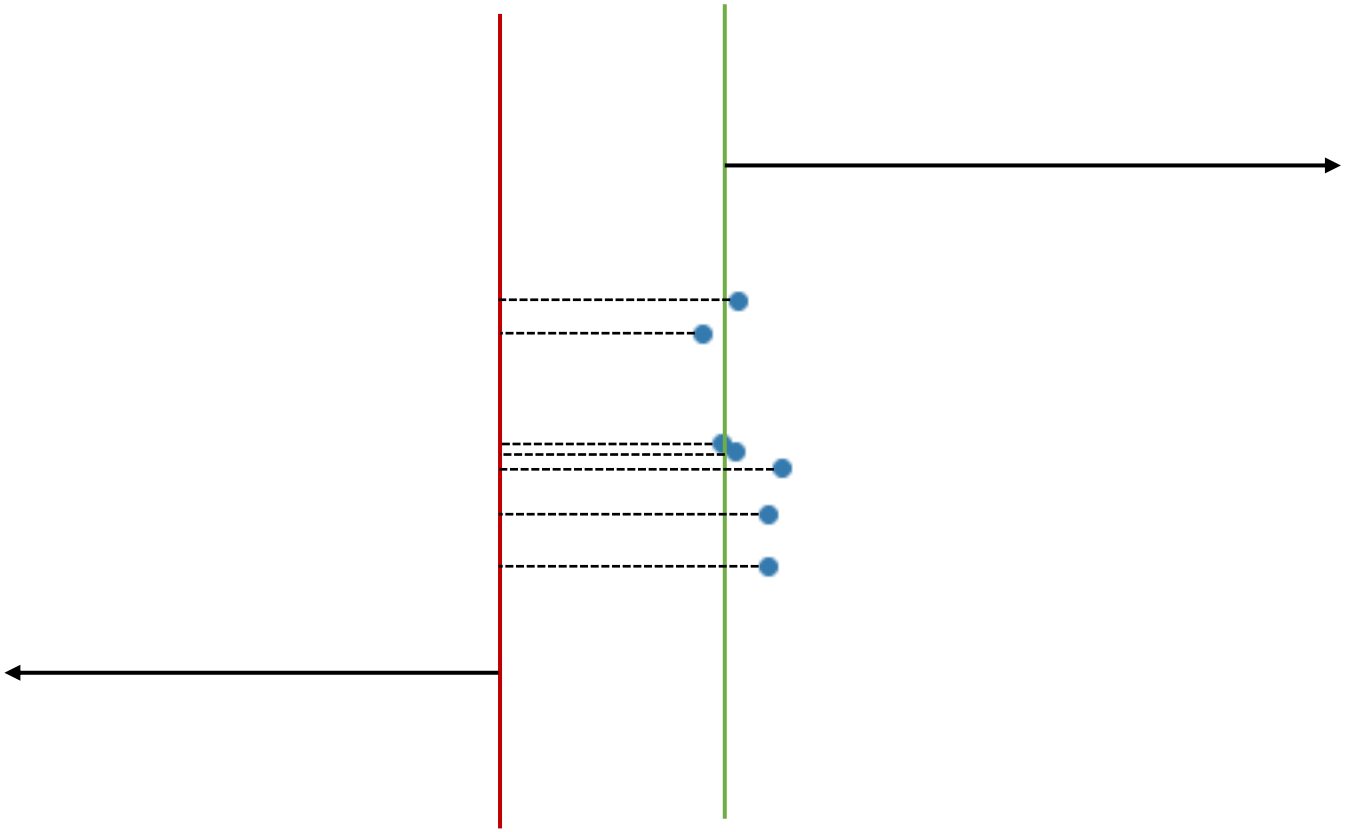


PCA



One dimensional subspace that maximizes the variance

PCA



One dimensional subspace that maximizes the variance

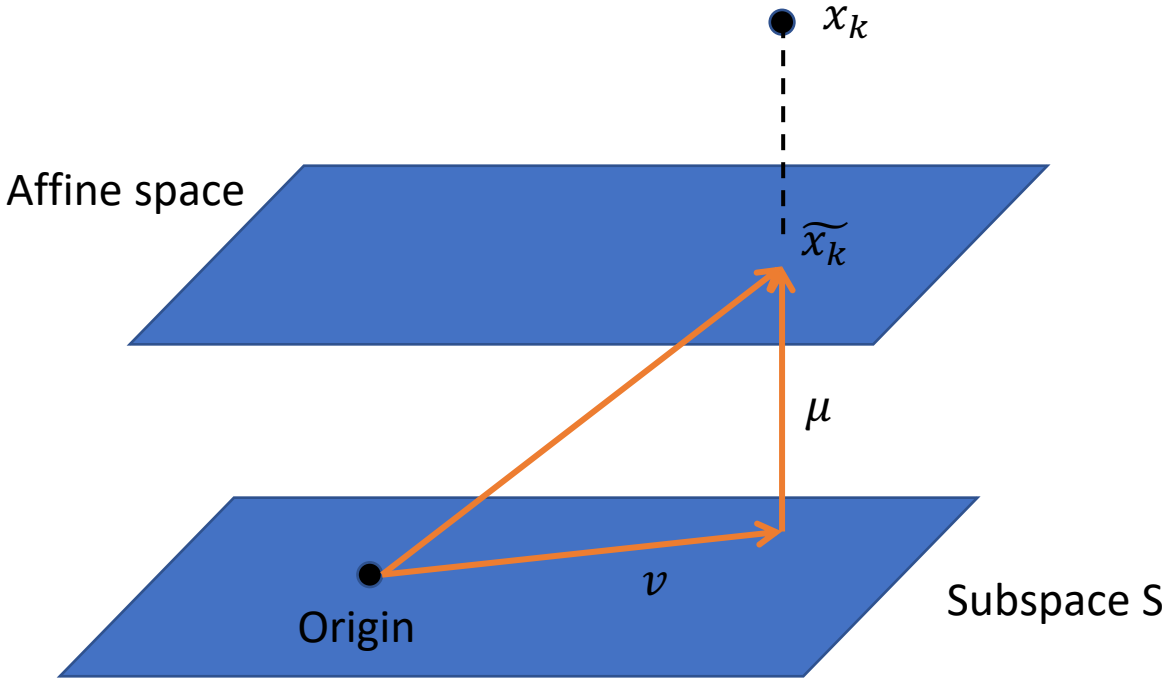
This also maximizes the variance. Why is this not the output of PCA?

Because we force our principal components to pass through origin which works perfectly only if the data is centered

PCA

1. PCA can also be thought of as finding a d-dimensional affine space such that the sum of $l2$ distance between the data points and their projection on the affine space is minimum. How to formulate this optimization mathematically?

PCA



Let $v_1, \ldots, v_d$ be an orthonormal basis of subspace S parallel to the affine space we are trying to find. Assume that μ is the distance between this subspace and the affine space we are trying to find.

Objective: $\min \sum_{k=1}^{n} \left|\left|x_k - \widetilde{x_k}\right|\right|^2$

Now, $\widetilde{x_k} = \mu + v$, where $v \in S$

Let $\mathrm{v} = \sum_{i=1}^{d} (\beta_k)_i v_i = V\beta_k$ (V is a matrix with v_i as the i^{th} column)

Thus, the objective function is: $\min \sum_{k=1}^{n} \left|\left|x_k - \mu - V\beta_k\right|\right|^2$

And this minimization is carried to find the affine space. Thus, we minimize the function over $\mu \in \mathbb{R}^p, \mathrm{V} \in \mathbb{R}^{p \times d}, \beta_k' s \in \mathbb{R}^d$

PCA


Critic cat
is not impressed by the artist's utter lack of dimensionality
ICANHASCHEEZBURGER.COM