Machine Learning (CSE 446): Principal Component Analysis and The Singular Value Decomposition

Sham M Kakade

University of Washington
cse446-staff@cs.washington.edu

PCA formulation 1: Dimension of Greatest Variance

Let $\mathbf{u}$ be the dimension of greatest variance, and (without loss of generality) let $\|\mathbf{u}\|_2^2 = 1$.

- $\mathbf{x}_i \cdot \mathbf{u}$ is the projection of the nth example onto $\mathbf{u}$.
- Since the mean of the data is $\mathbf{0}$, the mean of $\langle p_1, \ldots, p_N \rangle$ is also 0.
- This implies that the variance of $\langle p_1, \ldots, p_N \rangle$ is $\frac{1}{N}\sum_{i=1}^{N} p_i^2$.
- The $\mathbf{u}$ that gives the greatest variance, then, is:

$$\underset{\mathbf{u}}{\text{argmax}} \frac{1}{N}\sum_{i=1}^{N} (\mathbf{x}_i \cdot \mathbf{u})^2$$
$$\text{s.t. } \|\mathbf{u}\|_2^2 = 1$$

(This is PCA in one dimension!)

The optimization problem, in terms of matrices

- X is $N \times d$ data matrix.

$$\underset{\mathbf{u}}{\operatorname{argmax}} \ \|\mathbf{X}\mathbf{u}\|_2^2$$
$$\text{s.t. } \|\mathbf{u}\|_2^2 = 1$$

- The covariance matrix (assuming mean is subtracted):

$$\Sigma = \frac{1}{N} X^\top X = \frac{1}{N} \sum_{i=1}^{N} x_i x_i^\top$$

and, equivalently,

$$\underset{\mathbf{u}}{\operatorname{argmax}} \ \mathbf{u}^\top \Sigma \mathbf{u}$$
$$\text{s.t. } \|\mathbf{u}\|_2^2 = 1$$

Deriving the Solution

("relax". not as difficult as it looks

$$\max_{\mathbf{u}} \mathbf{u}^\top \Sigma \mathbf{u}, \text{ s.t. } \|\mathbf{u}\|_2^2 = 1$$

- The Lagrangian encoding: "relax" the constraint into the objective:

$$\max_{\mathbf{u}} \mathbf{u}^\top \Sigma \mathbf{u} - \lambda \|\mathbf{u}\|_2^2$$

- λ provides a 'soft' penalty
- (no we just do calculus). first derivatives with respect to $\mathbf{u}$): $2\Sigma\mathbf{u} - 2\lambda\mathbf{u}$
- Setting equal to $\mathbf{0}$ leads to: $\lambda\mathbf{u} = \Sigma\mathbf{u}$
- (maybe your recognize it?) this is the def. of an eigenvector ($\mathbf{u}$) and eigenvalue (λ) for the matrix Σ.
- We take the first (largest) eigenvalue.

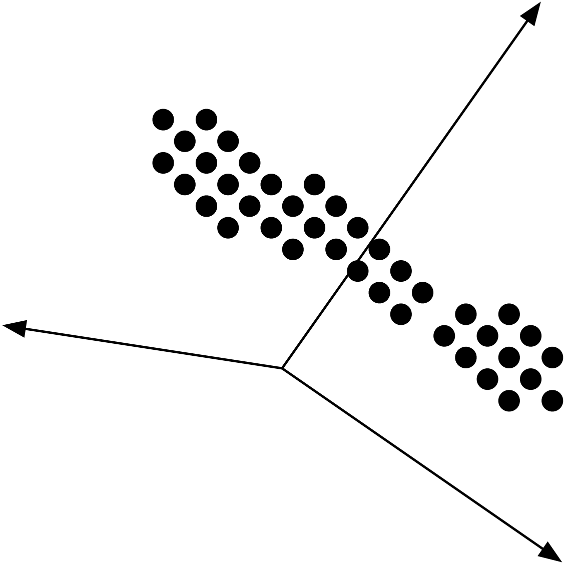
The solution.

- This is the 'eigenvalue' problem. We know the solution must satisfy:

$$\Sigma u = \lambda u$$

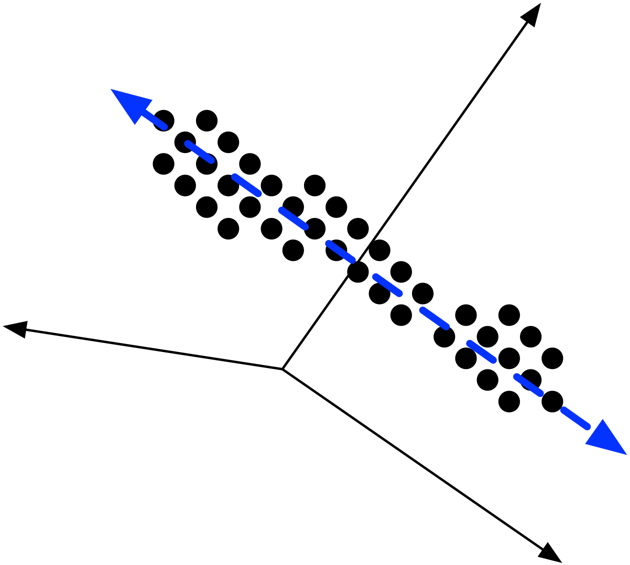
  for some λ.
- This means u is an eigenvector and λ is an eigenvalue of Σ.
- So to solve the optimization problem the eigenvector u corresponding to the largest eigenvalue λ.
- How do we find this?
  The Singular Value Decomposition.

Alternate View of PCA: Minimizing Reconstruction Error



Assume that the data are *centered*.
Find a line which minimizes the squared reconstruction error.

Alternate View of PCA: Minimizing Reconstruction Error



Assume that the data are *centered*.
Find a line which minimizes the squared reconstruction error.

Alternate View: Minimizing Reconstruction Error with K-dim subspace.

Equivalent ("dual") formulation of PCA: find an "orthonormal basis" $\mathbf{u_1}, \mathbf{u_2}, \dots \mathbf{u_K}$ which minimizes the total reconstruction error on the data:

$$\underset{\text{orthonormal basis:}\mathbf{u_1},\mathbf{u_2},\dots\mathbf{u_K}}{\operatorname{argmin}} \quad \frac{1}{N}\sum_i \|\mathbf{x}_i - \text{Proj}_{\mathbf{u_1},\dots\mathbf{u_K}}(\mathbf{x}_i)\|^2$$

Recall the projection of x onto K-orthonormal basis is:

$$\text{Proj}_{\mathbf{u_1},\dots\mathbf{u_K}}(\mathbf{x}) = \sum_{j=1}^{K} (\mathbf{u_i} \cdot \mathbf{x})\mathbf{u_i}$$

The SVD "simultaneously" finds all $\mathbf{u_1}, \mathbf{u_2}, \dots \mathbf{u_K}$

Principal Components Analysis: the algorithm

- Input: unlabeled data $\mathbf{X} = [\mathbf{x}_1 | \mathbf{x}_2 | \cdots | \mathbf{x}_N]^\top$; dimensionality $K < d$
- Output: K-dimensional "subspace".
- Algorithm:
  1. Compute the mean μ
  2. compute the **covariance matrix**:

$$\Sigma = \frac{1}{N} \sum_i (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^\top$$

  3. let $\langle \lambda_1, \ldots, \lambda_K \rangle$ be the top K eigenvalues of Σ and $\langle \mathbf{u}_1, \ldots, \mathbf{u}_K \rangle$ be the corresponding eigenvectors
- Let $\widetilde{\mathbf{U}} = [\mathbf{u}_1 | \mathbf{u} | \cdots | \mathbf{u}_K]$
  Return $\widetilde{\mathbf{U}}$

You can read about many algorithms for finding eigendecompositions of a matrix.

Projection and Reconstruction: the one dimensional case

- Take out mean μ:
- Find the "top" eigenvector u of the covariance matrix.
- What are your projections?
- What are your reconstructions, $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1 | \widehat{\mathbf{x}}_2 | \cdots | \widehat{\mathbf{x}}_N]^\top$?
- What is is your reconstruction error?

$$\frac{1}{N}\sum_i (\mathbf{x}_i - \widehat{\mathbf{x}}_i)^2 = ??$$

The singular value decomposition

- Let M be a symmetric matrix.
  SVDs also work for asymmetric matrices, with a slightly modified thm.
- SVD theorem: there exists a decomposition of the following form:

$$M = UDU^\top$$

  where D is a diagonal matrix and U is an orthogonal matrix (i.e. the columns of U are unit length and orthogonal to each other).
- The columns of U are eigenvectors of M.
- For PCA, you will take Σ to be M.