Kernel PCA

-Introduction to Data Science-

Lecturer: Darren Homrighausen, PhD

Principal Components Analysis (PCA)

Reminder:

Principal components analysis (PCA) is a dimension reduction technique

It estimates a new coordinate axis for the data that

- maximizes the variance of linear combinations
- minimizes ℓ_2 distortions

(These are equivalent)

PCA

In either case, we can compute it easily via the SVD

$$\mathbb{X} - \overline{\mathbb{X}} = (I - M)\mathbb{X} = UDV^\top$$

where $M = 11^\top/n$

(Note that sometimes the columns of $\mathbb{X}$ are centered and scaled)

Now, the

- PCA scores are found in the matrix UD

  (These are the of the observations in the PCs)

- PCA loadings are found in the matrix V

  (These are the coordinates of the features in the PCs)

Kernel PCA

Kernel PCA

Recall: The matrix $\mathbb{X}\mathbb{X}^\top$ is of the inner products $\langle X, X' \rangle$

Also, as $(I - M)\mathbb{X} = UDV^\top$, then

$$(I - M)\mathbb{X}\mathbb{X}^\top(I - M) = UD^2U^\top$$

Conclusion: The PCA scores can be computed via $\mathbb{X}\mathbb{X}^\top$

Kernel PCA

Now, to kernelize PCA, we need only replace:

$$\mathbb{X}\mathbb{X}^\top \to \mathbb{K}$$

where

$$\mathbb{K} = [k(X_i, X_{i'})]_{1 \le i, i' \le n}$$

and then write the decomposition

$$(I - M)\mathbb{X}\mathbb{X}^\top(I - M) \to (I - M)\mathbb{K}(I - M) = \tilde{U}\tilde{D}^2\tilde{U}^\top$$

(These $\tilde{U}$, $\tilde{D}$ are just used to indicate they are different from $(I - M)\mathbb{X}\mathbb{X}^\top(I - M) = UD^2U^\top$)

Kernel PCA

This approach still finds hyperplanes for dimension reduction

However, these hyperplanes are in a transformed space

$\rightarrow$ nonlinear dimension reduction in the original space

Example: We could make a classifier by doing:

1. Specify a k, which produces a $\mathbb{K}$ out of $\mathbb{X}$
2. Find $(I - M)\mathbb{K}(I - M) = \tilde{U}\tilde{D}^2\tilde{U}^\top$
3. Form feature matrix $\tilde{W} = \tilde{U}\tilde{D}[, 1 : m]$ for $m < \min\{n, p\}$
4. Now, train g on $\{(\tilde{W}_1, Y_1), \ldots, (\tilde{W}_n, Y_n)\}$

(This is a semi-supervised method)

Kernel PCA and beyond

The kernel PCA idea has strong connections to the nonlinear dimension reduction or manifold learning field:

- Laplacian eigenmaps
- diffusion maps
- locally linear embeddings
- principal curves