Feature Learning

Finding the diamonds in the rough

High Dimensionality

Higher Dimensionality

More Data

Feature 1	Feature 2	Feature 3	Feature 4
Value 1, 1	Value 1, 2	Value 1, 3	Value 1, 4
Value 2, 1	Value 2, 2	Value 2, 3	Value 2, 4
Value 3, 1	Value 3, 2	Value 3, 3	Value 3, 4
Value 4, 1	Value 4, 2	Value 4, 3	Value 4, 4
Value 5, 1	Value 5, 2	Value 5, 3	Value 5, 4
Value 6, 1	Value 6, 2	Value 6, 3	Value 6, 4

Feature Extraction

$$\min_{\mathbf{U},\mathbf{V}} ||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2,$$
$$s.t.\ \mathbf{U}^T\mathbf{U} = \mathbf{I}.$$

where **X** is our $d \times n$ data matrix, **U** is a $d \times r$ matrix of the principal components, and **V** is an $n \times r$ matrix of the data in the new principal components' space.

PCA's Goals

- Minimize reconstruction error
- Maximize variance along each principal component

PCA Goal

$$\min_{\mathbf{U},\mathbf{V}} ||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2,$$
$$s.t.\ \mathbf{U}^T\mathbf{U} = \mathbf{I}.$$

$$\mathbf{U}\mathbf{V}^T \approx \mathbf{X}.$$

PCA Goal

$$\begin{aligned} \mathbf{U}\mathbf{V}^T &= \mathbf{X}, \\ \mathbf{U}^T\mathbf{U}\mathbf{V}^T &= \mathbf{U}^T\mathbf{X}, \\ \mathbf{V}^T &= \mathbf{U}^T\mathbf{X}. \end{aligned}$$

PCA Alternative Definition

$$\min_{\mathbf{W}} ||\mathbf{X} - \mathbf{W}\mathbf{W}^T\mathbf{X}||_F^2,$$
$$s.t.\ \mathbf{W}^T\mathbf{W} = \mathbf{I}.$$

Covariance

$$\mathbf{C} = \frac{\mathbf{X}^T\mathbf{X}}{n-1}.$$

Covariance Relationship to Frobenius

$$\mathbf{C} = \frac{\mathbf{X}^T\mathbf{X}}{n-1}.$$

$$||\mathbf{X}||_F^2 = \mathbf{tr}(\mathbf{X}^T\mathbf{X}).$$

$$\mathbf{C} \propto ||\mathbf{X}||_F^2.$$

PCA Objective Reformulation (1)

$$||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 = \mathbf{tr}((\mathbf{X} - \mathbf{U}\mathbf{V}^T)(\mathbf{X} - \mathbf{U}\mathbf{V}^T)^T),$$

Using $||\mathbf{A}||_F^2 = \mathbf{tr}(\mathbf{A}^T\mathbf{A})$.

PCA Objective Reformulation (2)

$$\begin{aligned}||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}((\mathbf{X} - \mathbf{U}\mathbf{V}^T)(\mathbf{X} - \mathbf{U}\mathbf{V}^T)^T), \\ &= \mathbf{tr}((\mathbf{X} - \mathbf{U}\mathbf{V}^T)(\mathbf{X}^T - (\mathbf{U}\mathbf{V}^T)^T))\end{aligned}$$

Using $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.

PCA Objective Reformulation (3)

$$\begin{aligned} ||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}((\mathbf{X} - \mathbf{U}\mathbf{V}^T)(\mathbf{X}^T - (\mathbf{U}\mathbf{V}^T)^T)) \\ &= \mathbf{tr}((\mathbf{X} - \mathbf{U}\mathbf{V}^T)(\mathbf{X}^T - \mathbf{V}\mathbf{U}^T)) \end{aligned}$$

Using $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

PCA Objective Reformulation (4)

$$\begin{aligned}||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}((\mathbf{X} - \mathbf{U}\mathbf{V}^T)(\mathbf{X}^T - \mathbf{V}\mathbf{U}^T)) \\ &= \mathbf{tr}(\mathbf{X}\mathbf{X}^T - \mathbf{X}\mathbf{V}\mathbf{U}^T - \mathbf{U}\mathbf{V}^T\mathbf{X}^T + \mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{U}^T)\end{aligned}$$

Using $(a - b)(c - d) = ac - ad - bc + bd$.

PCA Objective Reformulation (5)

$$
\begin{aligned}
||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}(\mathbf{X}\mathbf{X}^T - \mathbf{X}\mathbf{V}\mathbf{U}^T - \mathbf{U}\mathbf{V}^T\mathbf{X}^T + \mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{U}^T) \\
&= \mathbf{tr}(-\mathbf{X}\mathbf{V}\mathbf{U}^T - \mathbf{U}\mathbf{V}^T\mathbf{X}^T + \mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{U}^T)
\end{aligned}
$$

Ignoring constants in objective function.

PCA Objective Reformulation (6)

$$
\begin{aligned}
||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}(-\mathbf{X}\mathbf{V}\mathbf{U}^T - \mathbf{U}\mathbf{V}^T\mathbf{X}^T + \mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{U}^T) \\
&= \mathbf{tr}(-\mathbf{X}\mathbf{X}^T\mathbf{U}\mathbf{U}^T - \mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T + \mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}\mathbf{U}^T)
\end{aligned}
$$

Using $\mathbf{V}^T = \mathbf{U}^T\mathbf{X}, \mathbf{V} = \mathbf{X}^T\mathbf{U}$.

PCA Objective Reformulation (7)

$$
\begin{aligned}
||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}(-\mathbf{X}\mathbf{X}^T\mathbf{U}\mathbf{U}^T - \mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T + \mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}\mathbf{U}^T) \\
&= \mathbf{tr}(-\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} - \mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} + \mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}\mathbf{U}^T\mathbf{U})
\end{aligned}
$$

Using $\mathbf{tr}(\mathbf{AB}) = \mathbf{tr}(\mathbf{BA})$.

PCA Objective Reformulation (8)

$$\begin{aligned}||\mathbf{X}-\mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}(-\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}-\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}+\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}\mathbf{U}^T\mathbf{U})\\ &= \mathbf{tr}(-\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}-\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}+\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U})\end{aligned}$$

Using $\mathbf{U}^T\mathbf{U}=\mathbf{I}$.

PCA Objective Reformulation (9)

$$
\begin{aligned}
||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}(-\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} - \mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} + \mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}) \\
&= \mathbf{tr}(-\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U})
\end{aligned}
$$

Using summation.

PCA Objective Reformulation (10)

$$\begin{aligned} ||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= \mathbf{tr}(-\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}) \\ &= -\mathbf{tr}(\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}) \end{aligned}$$

Using $\mathbf{tr}(-\mathbf{A}) = -\mathbf{tr}(\mathbf{A})$.

PCA Objective Reformulation (11)

$$\begin{aligned} \|\mathbf{X} - \mathbf{U}\mathbf{V}^T\|_F^2 &= -\mathbf{tr}(\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}) \\ &= -\mathbf{tr}(\mathbf{U}^T\mathbf{X}(\mathbf{U}^T\mathbf{X})^T) \end{aligned}$$

Using $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

PCA Objective Reformulation (12)

$$
\begin{aligned}
||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 &= -\mathbf{tr}(\mathbf{U}^T\mathbf{X}(\mathbf{U}^T\mathbf{X})^T) \\
&= -||\mathbf{U}^T\mathbf{X}||_F^2
\end{aligned}
$$

Using $||\mathbf{A}||_F^2 = \mathbf{tr}(\mathbf{A}^T\mathbf{A})$.

PCA Objective

$$
\begin{aligned}
&\min_{\mathbf{U},\mathbf{V}} \; ||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2, && s.t.\ \mathbf{U}^T\mathbf{U} = \mathbf{I}. \\
=&\min_{\mathbf{U}} \; -||\mathbf{U}^T\mathbf{X}||_F^2, && s.t.\ \mathbf{U}^T\mathbf{U} = \mathbf{I}. \\
=&\max_{\mathbf{U}} \; ||\mathbf{U}^T\mathbf{X}||_F^2, && s.t.\ \mathbf{U}^T\mathbf{U} = \mathbf{I}. \\
\approx&\max_{\mathbf{U}} \; cov(\mathbf{U}^T\mathbf{X}), && s.t.\ \mathbf{U}^T\mathbf{U} = \mathbf{I}.
\end{aligned}
$$

Singular Value Decomposition

$$\mathbf{X} = \mathbf{USV}^T,$$

where $\mathbf{X}$ is a $d \times n$ matrix, $\mathbf{U}$ is a unitary $d \times d$ matrix, $\mathbf{S}$ is a diagonal matrix of singular values, which are the square root of the respective eigenvalues, and $\mathbf{V}$ is a unitary $n \times n$ matrix.

Covariance Matrix SVD Representation

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T.$$

$$\begin{aligned}
\mathbf{C} &= \frac{\mathbf{X}^T\mathbf{X}}{n-1}, \\
&= \frac{\mathbf{V}\mathbf{S}\mathbf{U}^T\mathbf{U}\mathbf{S}\mathbf{V}^T}{n-1}, \\
&= \mathbf{V}\frac{\mathbf{S}^2}{n-1}\mathbf{V}^T.
\end{aligned}$$

Eigenvalue Decomposition

$$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T,$$

where $\mathbf{Q}$ is a matrix of eigenvector columns and Λ is a diagonal of the eigenvalues of $\mathbf{A}$.

Covariance Eigenvalue Decomposition

$$\mathbf{C} = \mathbf{V}\Lambda\mathbf{V}^T,$$

where $\mathbf{V}$ contains the eigenvector columns and Λ is the diagonal of eigenvalues of $\mathbf{C}$.

Singular Value, Eigenvalue Relationship

$$\frac{\mathbf{S}^2}{n-1} = \Lambda.$$

Solving PCA

We can calculate the principal components of a matrix by calculating the eigenvectors of its covariance matrix with the largest r eigenvalues,

$$\mathbf{U} = eig(\mathbf{C}, r). \tag{1}$$

Demo

Kernel PCA

Allows for non-linear transformations using the kernel trick.

$$\mathbf{C} = \phi(\mathbf{X})^T \phi(\mathbf{X}).$$

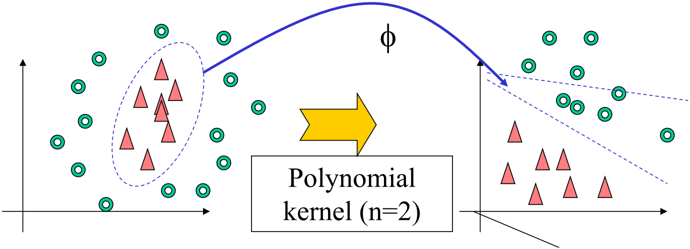
C is an $m \times m$ matrix.

Polynomial Kernel

$$\underbrace{K(\mathbf{x}_i, \mathbf{x_j})}_{\textit{Polynomial}} = (\mathbf{x}_i^T \mathbf{x}_j + 1)^d,$$

with degree d.

Polynomial Kernel

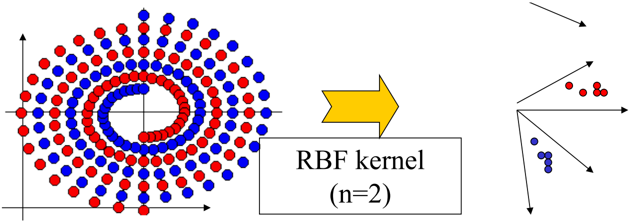


Radial Basis Function Kernel

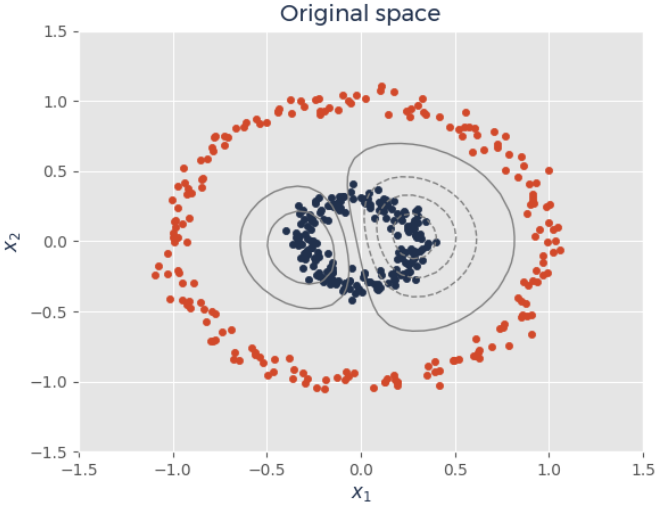
$$\underset{RBF}{K}(\mathbf{x}_i, \mathbf{x_j}) = \exp(\frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}),$$

with width σ.

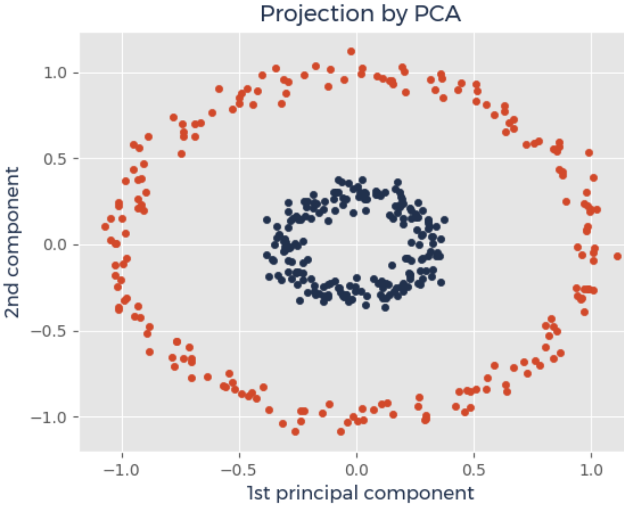
Radial Basis Function Kernel



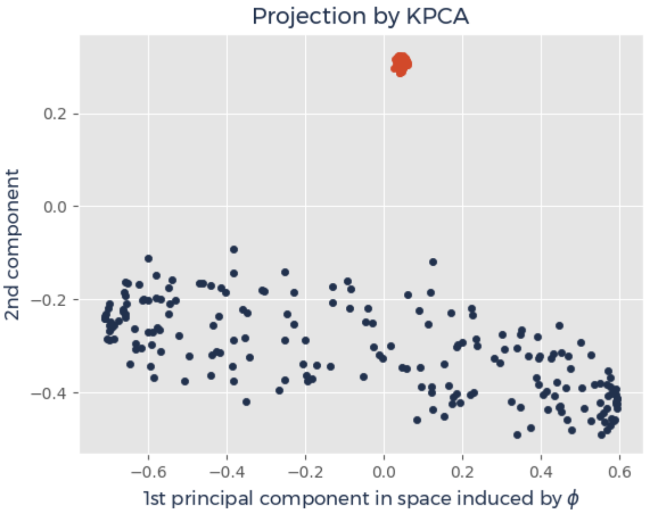
Kernel PCA Example



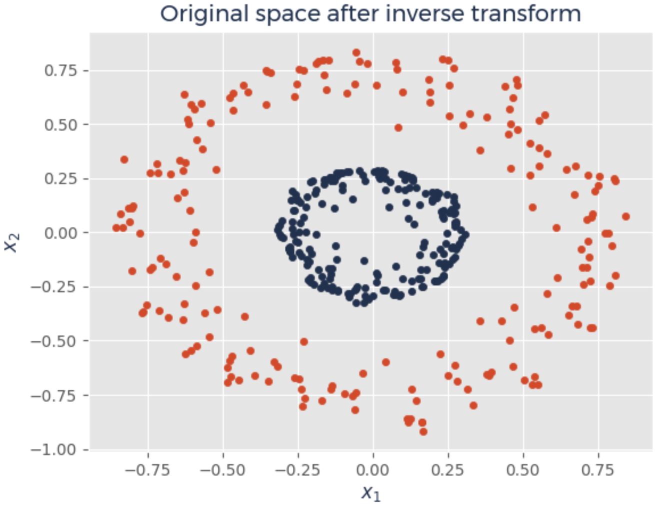
Kernel PCA Example



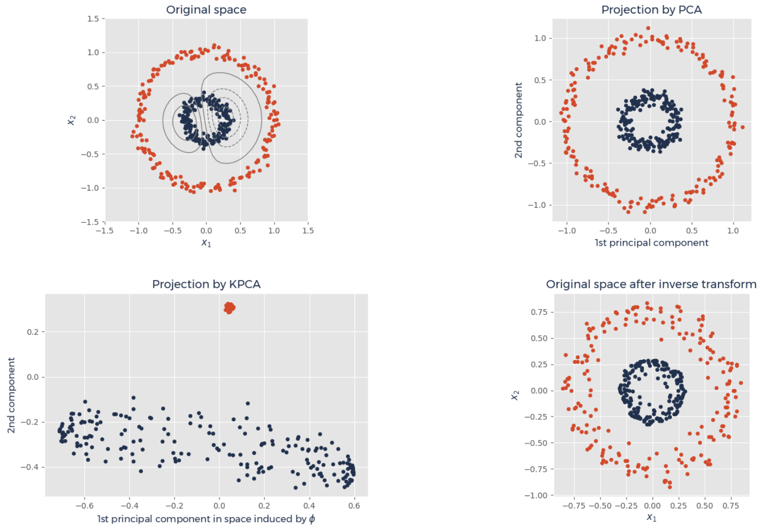
Kernel PCA Example



Kernel PCA Example



Kernel PCA Example



Questions

These slides are designed for educational purposes, specifically the CSCI-470 Introduction to Machine Learning course at the Colorado School of Mines as part of the Department of Computer Science.

Some content in these slides are obtained from external sources and may be copyright sensitive. Copyright and all rights therein are retained by the respective authors or by other copyright holders. Distributing or reposting the whole or part of these slides not for academic use is HIGHLY prohibited, unless explicit permission from all copyright holders is granted.