# ECE M146 Introduction to Machine Learning

Prof. Lara Dolecek

ECE Department, UCLA

# Today’s Lecture

Recap:

- Unsupervised Learning
- K-means algorithm for clustering

New topic:

- PCA algorithm for dimensionality reduction
- Linear algebra: SVD, eigen values

# Today's Lecture

Recap:

- Unsupervised Learning
- K-means algorithm for clustering

New topic:

- PCA algorithm for dimensionality reduction
- Linear algebra: SVD, eigen values

# Recap: Unsupervised learning

- In the previous lecture, we introduced the notion of unsupervised learning.

- This set-up is described by un-labeled data.

- The goal is to organize or represent this data.

# Clustering

- Clustering = group unlabeled data into clusters.
- Use prototypes as cluster representatives.
- This is a form of lossy data compression.

# K-means algorithm

- K-means algorithm is an iterative algorithm that iterates between two steps:

1) Cluster Assignment (set indicators given the prototypes)
2) Refitting (set prototypes given the indicators)

# Today's Lecture

Recap:

- Unsupervised Learning
- K-means algorithm for clustering

New topic:

- PCA algorithm for dimensionality reduction
- Linear algebra: SVD, eigen values

# Dimensionality reduction

- In clustering, we used prototypes to represent individual clusters (all data points assigned to that cluster).
- Prototypes are of the same dimension as the data points.

- Another way to represent data efficiently is to **project all data points onto a lower dimensional space.**

- Applications are also in lossy data compression, visualization and also feature extraction.

# Dimensionality reduction

- Set-up: Suppose data is provided in some D-dimensional space, but it can be well explained in an M-dimensional subspace for M<D.

- Example: D = 2 and M = 1.

- Goal: find this subspace.

# How to find the appropriate subspace of dimension M ?

- Some projections:

- We want to look at the projection with the **highest sample variance**, because this will be the most informative choice.

# First, we need some linear algebra

Notion of a **basis**:

- For a vector space V, a set of vectors $\{v_1, v_2,…,v_D\}$ forms a basis if:
- Vectors $v_1, v_2,…,v_D$ are linearly independent:
- Any vector x in SD is uniquely expressed in terms of $\{v_1, v_2,…,v_D\}$ as:
- Vector space V must then be of dimension D.

# Basis set is not unique

- Here are some examples in 3-D.
- Example 1:
- Example 2:
- Example 3:

# Orthonormal basis

- It is also of interest to consider a basis that has vectors that are orthogonal to each other and each has unit norm.

- Mathematically:
- Cond. 1:
- Cond. 2:
- This basis is then called **orthonormal basis**.

- Which of the preceding examples constitute orthogonal basis, orthonormal basis, neither ?

# Basis set is not unique

- Here are some examples in 3-D.
- Example 1:
- Example 2:
- Example 3:

# Matrix format

- Can organize vectors of an orthonormal basis into a matrix:
- Useful property of this matrix:

# Back to our set-up: projection with max. sample variance

- Suppose we have N data points, in a D-dimensional space: $x_1, x_2, \ldots, x_N$.

- Suppose we wish to project the data onto the most informative dimension. Let $u_1$ be a vector in that dimension.

- Then, these are our projections:

# Sample mean and its projection

- Sample mean

- By linearity, projected sample mean is the same as the projection of the sample mean.

# Sample variance in the projection dimension

- Consider the following:

- Write it in the matrix format:

# Math verification

# Maximization of the sample variance in the projection dimension

- The goal of finding the dimension along which the sample variance is maximized is then equivalent to maximizing $u_1^T \times S \times u_1$ ?

- Almost.
- Note that without further restriction, we could pick an arbitrary dimension and vector $u_1$ of huge magnitude and artificially maximize this product.
- To make this meaningful, consider only vectors of unit norm:

# Mathematical formulation

- What we then have is:

- This is a constrained optimization problem.
- Where did we study constrained optimization problems before ?

SVM

- How did we solve them then ?

Convert to unconstrained optimization and take derivative of the Lagrangian.

# Unconstrained optimization

- Mathematical formulation:

- Set the derivative to zero.

# Math, continued

# Eigenvalues and eigenvectors

# Upshot

- For any eigen vector, we get a stationary solution.
- Matrix S being a covariance matrix is positive semi-definite, so all its eigenvalues are non-negative, and all its eigenvectors are real.
- So we are looking for the eigenvector that corresponds to the largest eigenvalue!

# Arrived at PCA

- This is the dimension projection.
- This vector $u_1$ is called the **first principal component**.
- Hence the name: **principal component analysis.**

# General version of PCA

- We can generalize this projection idea to more than one dimension.
- Idea: express S as $S = X^T * X$ (view the matrix X as the centered version of data)

# General version of PCA

- Decompose X as follows:

- This is known as the singular value decomposition (SVD)

# Math analysis

- Write S as

# General PCA:

- Write S as

- SVD is a powerful matrix decomposition tool. We just showed how to use it in PCA.