Linear Algebra & Convex Optimization – Lecture 2

From Last Lecture

- Vector Scaling, Addition & Subtraction
- Data Representation as Vectors
- Vector Inner Product

Outline

- Vector Norms
- Vector Distances Use-case – Clustering

Textbook: Introduction to Applied Linear Algebra, S. Boyd: Chapters 4,5.

Vector Norms

Euclidean norm (L2) of n-dimensional vector x :

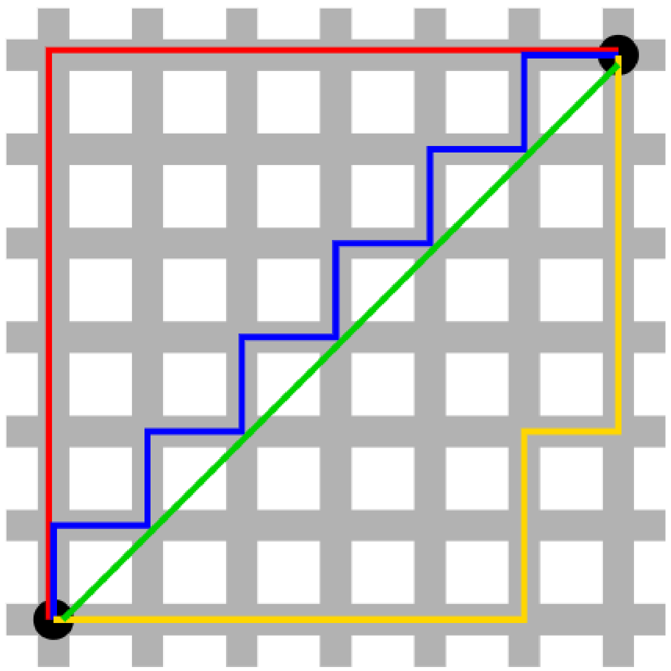
$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \qquad \|x\|_2 = \|x\| = \sqrt{x^T x}$$

L1 - Norm:

$$||x||_1 = \sum_i |x_i| = |x_1| + |x_2| + \ldots + |x_i|$$

Norm of Sum:

$$\|x+y\|^2 = (x+y)^T(x+y) = x^T x + x^T y + y^T x + y^T y$$
$$= \|x\|^2 + 2x^T y + \|y\|^2$$



L1-Norm vs L2- Norm

Poll: We have two vectors a_1, b_1 which has noisy components . In this case which norm is more reliable to compute the distance between the vectors, L1 or L2?

L1 for larger noise, L2 for smaller noise (< 1)

Distances : Examples

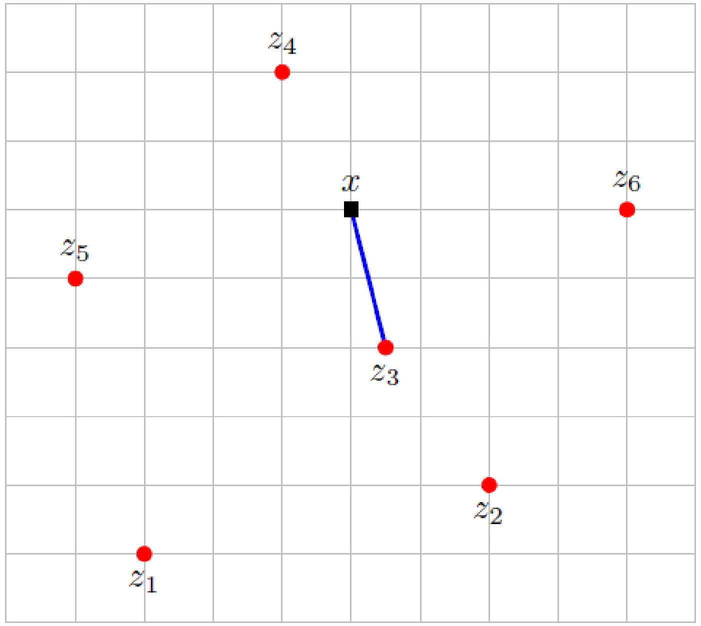
- **Nearest neighbor:**

$z_1, \ldots, z_m$ is a collection of m n-vectors

x is another n-vector

z_j is the *nearest neighbor* of x if :

$$\|x - z_j\| \leq \|x - z_i\|, \quad i = 1, \ldots, m$$



- **Document Distances:**

Document represented as vectors : histogram of words

Difference of vector norms are distances between documents

Pairwise Distances between various Wikipedia articles

	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	0.095	0.130	0.153	0.170
Memorial Day	0.095	0	0.122	0.147	0.164
Academy A.	0.130	0.122	0	0.108	0.164
Golden Globe A.	0.153	0.147	0.108	0	0.181
Super Bowl	0.170	0.164	0.164	0.181	0

Standardization (Vector Dimension)

$$\|x-y\|^2 = (x_1-y_1)^2 + \cdots + (x_n-y_n)^2$$

Comparison of two time –series data

$$\mathbf{avg}(x) = \mu = \mathbf{1}^T x/n$$

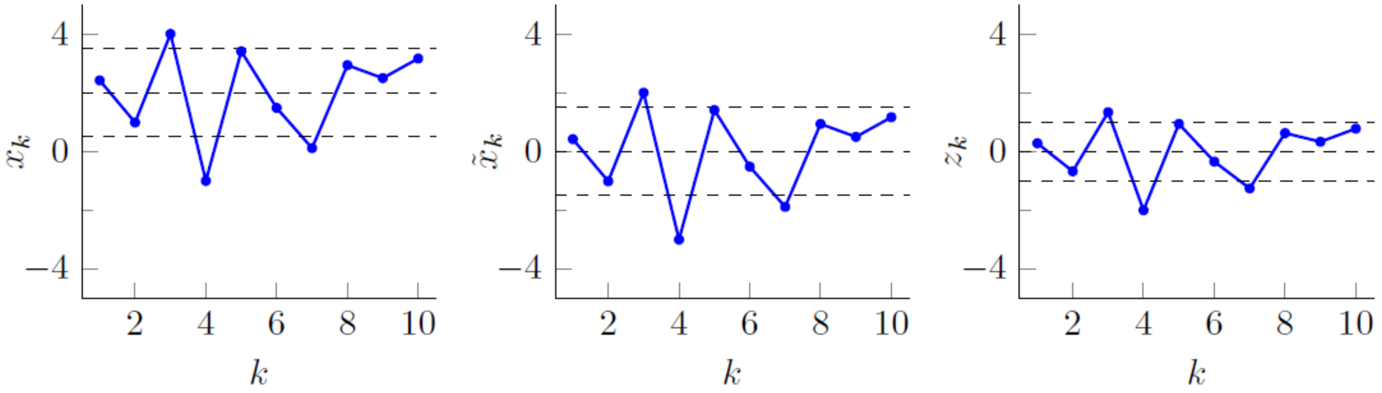
$$\mathbf{std}(x) = \sigma = \|x - \mu\mathbf{1}\|/\sqrt{n}$$

Standard Deviation : How entries of vector deviate from their mean value.

$$z = \frac{1}{\mathbf{std}(x)}(x - \mathbf{avg}(x)\mathbf{1})$$

Standardization : Mean = 0, Deviation = 1

Vector Standardization:



Vector Standardization: Useful for comparing 1D series data

Standardization (Feature Dimension)

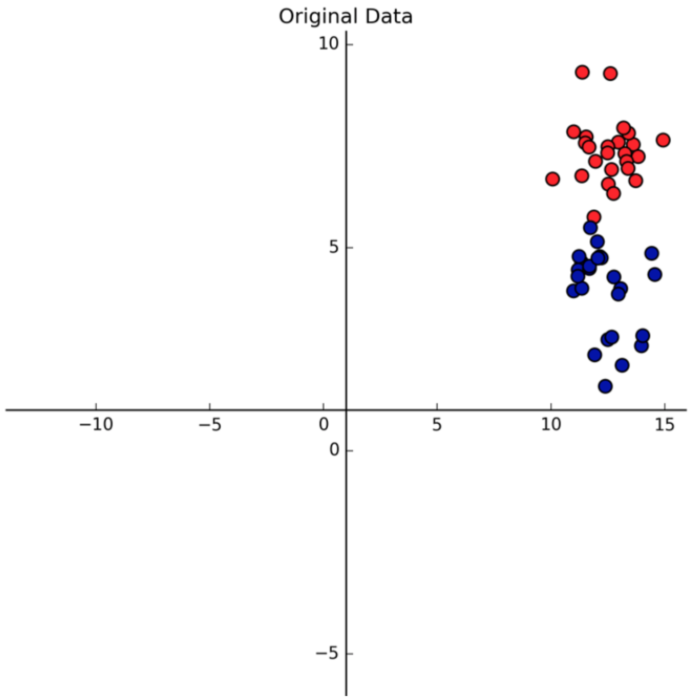
$$Z(:,\mathrm{i}) = \frac{1}{\textbf{std}\ (X(:,i))}[X(:,\mathrm{i}) - \textbf{avg}(\ X(:,\mathrm{i}))]$$

Standardization is done along feature dimension

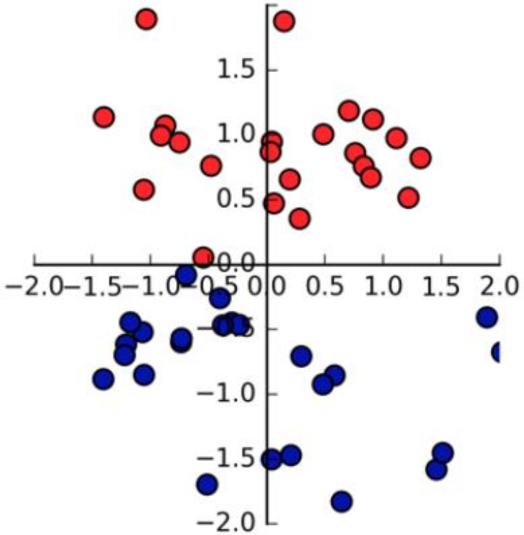
Useful in ML applications

Data Standardization:

Person Name	**Salary**	**Year of Experience**	**Expected Position Level**
Aman	100000	10	2
Abhinav	78000	7	4
Ashutosh	32000	5	8
Dishi	55000	6	7
Abhishek	1200000	8	3



Before Data Standardization



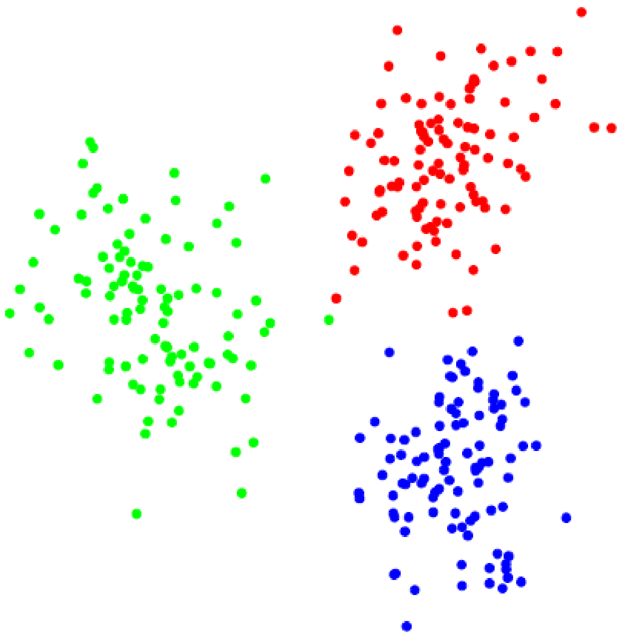
After Data Standardization

Clustering

Objective: Given a set of N n–vectors, $x_1, x_2, \ldots, x_N$ find k groups/clusters where $k << N$

Before Clustering

$k = 3$
$n = 2$
$N = 100$



After Clustering

Example Applications:

1. Document Topic Discovery
   (x_i : word histograms)

2. Patient Clustering
   (x_i : patient details)

3. Customer Segmentation
   (x_i : purchase quantities for various items)

Clustering Formulation

Assume K " Group Representative/Cluster Centroids" n –vectors :
$z_1, z_2, z_3, \ldots z_\mathrm{k}$

[$z_1, z_2, z_3, \ldots z_\mathrm{k}$ need not be from $x_1, x_2, \ldots, x_N$]

$c_1, c_2, \ldots c_N$, $c_i \in \{1, 2, \cdots k\}$ are group assignment variables

x_i is in group $j = c_i$ $\Rightarrow$ z_{c_i} is the representative vector associated with x_i.

Clustering Objective

Objective:

To seek a choice of group assignments , $c_1, c_2, \dots c_N$, $c_i \in \{1, 2, \cdots k\}$, that minimizes the objective

$$J^{\text{clust}} = \left(\|x_1 - z_{c_1}\|^2 + \cdots + \|x_N - z_{c_N}\|^2\right)/N$$

Solution:

1. With representatives $(z_1, z_2, z_3, \dots z_{\mathrm{k}})$ fixed, find the best cluster assignments $(c_1, c_2, \dots c_N)$ that minimizes J^{clust}

2. With cluster assignments $(c_1, c_2, \dots c_N)$ fixed , find the best representatives $(z_1, z_2, z_3, \dots z_{\mathrm{k}})$ that minimizes J^{clust}

Best Clustering

Representatives ($z_1, z_2, z_3, \dots z_k$) fixed:

$$J^{\text{clust}} = \left(\|x_1 - z_{c_1}\|^2 + \cdots + \|x_N - z_{c_N}\|^2\right)/N$$

Minimize each $\|x_i - z_{c_i}\|^2$ term independently to minimize J^{clust}

Nearest representative to x_i will be chosen : If z_j is the nearest representative, $c_i = j$

$$\|x_i - z_{c_i}\| = \min_{j=1,\ldots,k} \|x_i - z_j\|,$$

$$J^{\text{clust}} = \left(\min_{j=1,\ldots,k} \|x_1 - z_j\|^2 + \cdots + \min_{j=1,\ldots,k} \|x_N - z_j\|^2\right)/N.$$

Cluster Assignments ($c_1, c_2, \dots c_N$) fixed: :

Rewrite $J^{\text{clust}} = J_1 + \cdots + J_k$ where , $J_j = (1/N) \sum_{i \in G_j} \|x_i - z_j\|^2$

$$z_j = (1/|G_j|) \sum_{i \in G_j} x_i$$

Poll:

Will the cluster representatives(z) change a different distance metric is used

yes

K- Means Algorithm

k-MEANS ALGORITHM

given a list of N vectors $x_1, \ldots, x_N$, and an initial list of k group representative vectors $z_1, \ldots, z_k$

repeat until convergence

1. *Partition the vectors into k groups.* For each vector $i = 1, \ldots, N$, assign x_i to the group associated with the nearest representative.
2. *Update representatives.* For each group $j = 1, \ldots, k$, set z_j to be the mean of the vectors in group j.

K- Means illustration

K – Means Illustration
K = 4

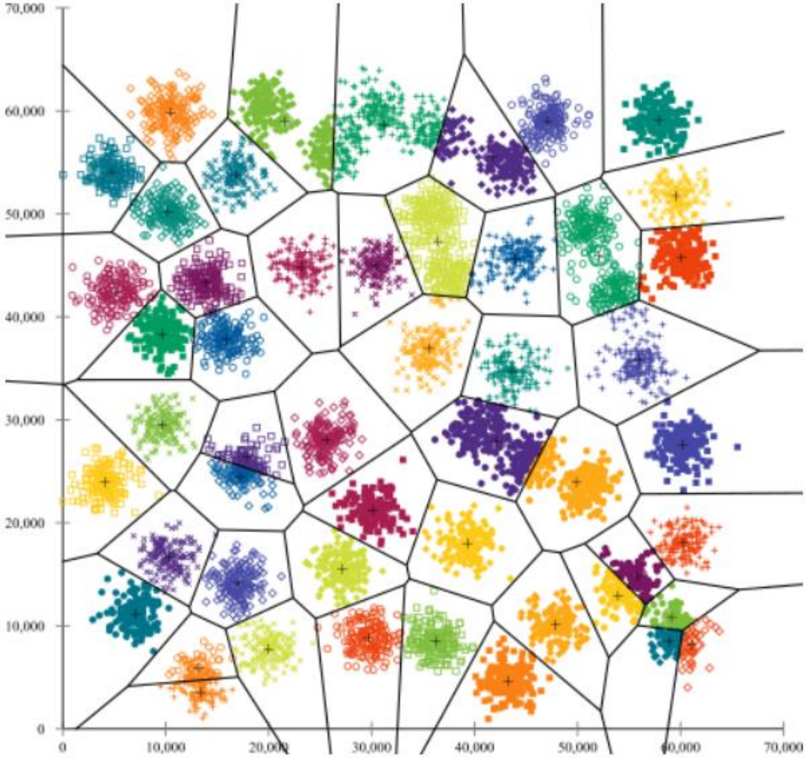
K-Means Challenges :

Challenges:

1. Clustering depends on the value of 'K' . Difficult to predict the ideal 'K' for many cases

2. Clustering varies with different initial points

3. With varying cluster sizes, optimal clustering is difficult to achieve



Initialization problem in K-Means with large number of clusters. (refer k-means++)