COL866: Foundations of Data Science

Ragesh Jaiswal, IITD

Singular vectors and Eigenvectors

- For a square matrix B, if $B\mathbf{x} = \lambda\mathbf{x}$, then $\mathbf{x}$ is said to be an *eigenvector* of B and λ the corresponding *eigenvalue*.
- For a rectangular $n \times d$ matrix A, let $B = A^T A$.
- Claim 1: The right singular vectors of A are the eigenvectors of B.
- Claim 2: The left singular vectors of A are the eigenvectors of AA^T.

- For a square matrix B, if $B\mathbf{x} = \lambda\mathbf{x}$, then $\mathbf{x}$ is said to be an *eigenvector* of B and λ the corresponding *eigenvalue*.
- For a rectangular $n \times d$ matrix A, let $B = A^T A$.
- Claim 1: The right singular vectors of A are the eigenvectors of B.
- Claim 2: The left singular vectors of A are the eigenvectors of AA^T.
- Claim 3: For every $\mathbf{x}, \mathbf{x}^T B\mathbf{x} \geq 0$. Such matrices are called *positive semidefinite*.
- Fact: Any positive semidefinite matrix can be decomposed into a product $A^T A$.
- So, *eigen decomposition* of any positive semidefine matrix may be obtained from SVD of A.

- The line through the origin that minimizes the sum of squared distances is defined by the first singular vector.
- Question: What if we drop the constraint that the line should pass through the origin? This general case is more relevant in data analysis.

- The line through the origin that minimizes the sum of squared distances is defined by the first singular vector.
- Question: What if we drop the constraint that the line should pass through the origin? This general case is more relevant in data analysis.

Theorem

The best fit line of a set of data points must pass through the centroid of the points.

Theorem

The best fit line of a set of data points must pass through the centroid of the points.

- We can generalize for higher dimensional objects.

Definition (Affine subspace)

An affince subspace is a subspace translated by a vector. So, it is a set of the form: $\{\mathbf{v}_0 + \sum_{i=1}^{k} c_i \mathbf{v}_i | c_1, ..., c_k \in \mathbb{R}\}$. Here $\mathbf{v}_0$ is the translation and $\mathbf{v}_1, ..., \mathbf{v}_k$ form an orthonormal basis of the subspace.

Theorem

The k-dimensional affine space that minimizes the sum of squared perpendicular distances to the data points must pass through the centroid of the points.

Theorem

The best fit line of a set of data points must pass through the centroid of the points.

Theorem

The k-dimensional affine space that minimizes the sum of squared perpendicular distances to the data points must pass through the centroid of the points.

- Given the above, the standard first step in data analysis is to center the data which means subtracting the centroid from all the data points (so that the origin is the centroid).

- Given the above, the standard first step in data analysis is to center the data which means subtracting the centroid from all the data points (so that the origin is the centroid).
- Usually, a next natural data analysis investigation on the centered data is to find the principle components of the data. That is, the directions along which the data has the maximum variation.
- Our discussion of Singular values and Singular vectors are already in tune with this idea. Note that the Singular vectors give the direction along which the data has the maximum variance.
- If we consider the projection Y of the data points X onto V_k (i.e., the subspace defined by the first k Singular vectors), then Y may have much smaller dimension than X and yet has most of the "content" of the original data X.
- This is a standard dimension reduction technique and is popularly known as Principal Component Analysis (PCA).

- Linear regression: Given $\mathbf{x}_1, ..., \mathbf{x}_m \in \mathbb{R}^d$ and $y_1, ..., y_m \in \mathbb{R}$ we want to find a vector $\mathbf{w}$ that minimizes:

$$\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w} - y_i) = ||A\mathbf{w} - \mathbf{y}||^2.$$

  Where A is a matrix with the i^{th} row as $\mathbf{x}_i^T$.
- The following theorem gives a very nice application of SVD.

Theorem

The vector $\mathbf{w}$ that minimizes $||A\mathbf{w} - \mathbf{y}||^2$ is $\mathbf{w} = A^\dagger \mathbf{y} = VD^\dagger U^T \mathbf{y}$ for $A = UDV^T$ and $D_{ii}^\dagger = 1/D_{ii}$ for all i.

Spectral Graph Theory: Eigenvalues and graph properties

Spectral Graph Theory

- Graphs are used to model pairwise relationship between objects. The objects are denoted by vertices and pairwise relationships are denoted by edges.
- Classical techniques have had a lot of success in solving various problems on graphs such as shortest path, minimum spanning tree etc.
- More recently, algebraic techniques have been successful for certain graph problems. Such techniques involve analysing the adjacency matrix of the graph.
- The relationship between the structural properties of a given graph and the algebraic properties of the underlying (appropriate variant of) adjacency matrix is key in such analysis.

Spectral Graph Theory

- Graphs are used to model pairwise relationship between objects. The objects are denoted by vertices and pairwise relationships are denoted by edges.
- Classical techniques have had a lot of success in solving various problems on graphs such as shortest path, minimum spanning tree etc.
- More recently, algebraic techniques have been successful for certain graph problems. Such techniques involve analysing the adjacency matrix of the graph.
- The relationship between the structural properties of a given graph and the algebraic properties of the underlying (appropriate variant of) adjacency matrix is key in such analysis.
- Spectral Graph Theory studies how eigenvalues of the adjacency matrix of a graph (an algebraic quantity) relates to combinatorial properties of the graph.

- We will study connectivity property of graphs.
- Question: What is the "gold standard" for connectivity for graphs?

- We will study connectivity property of graphs. In many applications, connectivity is of vital importance.
- Question: What is the "gold standard" for connectivity for graphs? Complete graph or a clique
- Consider a communication network setting. Adding edges has an associated cost. So, ideally we would like to have good connectivity with small number of total edges.
- What does it mean to have good connectivity?

- We will study connectivity property of graphs. In many applications, connectivity is of vital importance.
- Question: What is the "gold standard" for connectivity for graphs? Complete graph or a clique
- Consider a communication network setting. Adding edges has an associated cost. So, ideally we would like to have good connectivity with small number of total edges.
- What does it mean to have good connectivity?
- An intuitive notion is that removing a small fraction of edges of the graph should not disconnect large portions of the graph.
- Sparsity and edge expansion discussed next, are useful in formalizing the notion of connectivity.

- Given a graph $G = (V, E)$, a cut in the graph is a partition of vertices into sets S and $V - S$. It is represented as a tuple $(S, V - S)$.

Definition (Sparsity)

The sparsity of a cut $(S, V - S)$, denoted by $\sigma(S)$, is defined as $\sigma(S) = \frac{|V|^2}{|E|} \cdot \frac{|E(S, V-S)|}{|S| \cdot |V-S|}$.

- In other words, the sparsity of the cut $(S, V - S)$ is the fraction of edges across the cut divided by the fraction of vertex pairs across the partition (i.e., the fraction of edges in the ideal case of a complete graph).

- Given a graph $G = (V, E)$, a cut in the graph is a partition of vertices into sets S and $V - S$. It is represented as a tuple $(S, V - S)$.

Definition (Sparsity)

The sparsity of a cut $(S, V - S)$, denoted by $\sigma(S)$, is defined as $\sigma(S) = \frac{|V|^2}{|E|} \cdot \frac{|E(S,V-S)|}{|S|\cdot|V-S|}$.

- In other words, the sparsity of the cut $(S, V - S)$ is the fraction of edges across the cut divided by the fraction of vertex pairs across the partition (i.e., the fraction of edges in the ideal case of a complete graph).
- The sparsest cut problem is to find a non-trivial cut with minimal sparsity.

- Given a graph $G = (V, E)$, a cut in the graph is a partition of vertices into sets S and $V - S$. It is represented as a tuple $(S, V - S)$.

Definition (Sparsity)

The sparsity of a cut $(S, V - S)$, denoted by $\sigma(S)$, is defined as $\sigma(S) = \frac{|V|^2}{|E|} \cdot \frac{|E(S, V-S)|}{|S| \cdot |V-S|}$.

- The sparsest cut problem is to find a non-trivial cut with minimal sparsity.
- The sparsity of a graph $G = (V, E)$, denoted by $\sigma(G)$, is given by:

$$\sigma(G) = \min_{S \subseteq V, S \neq \emptyset, S \neq V} \sigma(S).$$

Definition (Sparsity)

The sparsity of a cut $(S, V - S)$, denoted by $\sigma(S)$, is defined as $\sigma(S) = \frac{|V|^2}{|E|} \cdot \frac{|E(S,V-S)|}{|S| \cdot |V-S|}$.

- The sparsity of a graph $G = (V, E)$, denoted by $\sigma(G)$, is given by:

$$\sigma(G) = \min_{S \subseteq V, S \neq \emptyset, S \neq V} \sigma(S).$$

- Note that if the graph is d-regular, then $\sigma(S) = \frac{|E(S,V-S)|}{\frac{d}{|V|} \cdot |S| \cdot |V-S|}$
- The edge expansion of a set of vertices $S \subseteq V$, denoted by $\phi(S)$, of a d-regular graph is given by:

$$\phi(S) = \frac{|E(S, V - S)|}{d \cdot |S|}$$

Definition (Sparsity)

The sparsity of a cut $(S, V - S)$, denoted by $\sigma(S)$, is defined as $\sigma(S) = \frac{|V|^2}{|E|} \cdot \frac{|E(S,V-S)|}{|S| \cdot |V-S|}$.

- The sparsity of a graph $G = (V, E)$, denoted by $\sigma(G)$, is given by:

$$\sigma(G) = \min_{S \subseteq V, S \neq \emptyset, S \neq V} \sigma(S).$$

- Note that if the graph is d-regular, then $\sigma(S) = \frac{|E(S,V-S)|}{\frac{d}{|V|} \cdot |S| \cdot |V-S|}$
- The edge expansion of a set of vertices $S \subseteq V$, denoted by $\phi(S)$, of a d-regular graph is given by: $\phi(S) = \frac{|E(S,V-S)|}{d \cdot |S|}$.
- The edge expansion of a d-regular graph is given by $\phi(G) = \min_{S:|S| \leq |V|/2} \phi(S)$.

- Note that if the graph is d-regular, then $\sigma(S) = \frac{|E(S, V-S)|}{\frac{d}{|V|} \cdot |S| \cdot |V-S|}$
- The edge expansion of a set of vertices $S \subseteq V$, denoted by $\phi(S)$, of a d-regular graph is given by: $\phi(S) = \frac{|E(S, V-S)|}{d \cdot |S|}$.
- The edge expansion of a d-regular graph is given by $\phi(G) = \min_{S:|S| \leq |V|/2} \phi(S)$.
- Claim 1: For every set S such that $|S| \leq |V|/2$, $\frac{1}{2}\sigma(S) \leq \phi(S) \leq \sigma(S)$.

- Note that if the graph is d-regular, then $\sigma(S) = \frac{|E(S, V-S)|}{\frac{d}{|V|} \cdot |S| \cdot |V-S|}$
- The edge expansion of a set of vertices $S \subseteq V$, denoted by $\phi(S)$, of a d-regular graph is given by: $\phi(S) = \frac{|E(S, V-S)|}{d \cdot |S|}$.
- The edge expansion of a d-regular graph is given by $\phi(G) = \min_{S : |S| \leq |V|/2} \phi(S)$.
- Claim 1: For every set S such that $|S| \leq |V|/2$, $\frac{1}{2}\sigma(S) \leq \phi(S) \leq \sigma(S)$.
- Claim 2: $\frac{1}{2}\sigma(G) \leq \phi(G) \leq \sigma(G)$.

- Note that if the graph is d-regular, then $\sigma(S) = \frac{|E(S,V-S)|}{\frac{d}{|V|}\cdot|S|\cdot|V-S|}$
- The edge expansion of a set of vertices $S \subseteq V$, denoted by $\phi(S)$, of a d-regular graph is given by: $\phi(S) = \frac{|E(S,V-S)|}{d\cdot|S|}$.
- The edge expansion of a d-regular graph is given by $\phi(G) = \min_{S:|S|\leq|V|/2} \phi(S)$.
- Claim 1: For every set S such that $|S| \leq |V|/2$, $\frac{1}{2}\sigma(S) \leq \phi(S) \leq \sigma(S)$.
- Claim 2: $\frac{1}{2}\sigma(G) \leq \phi(G) \leq \sigma(G)$.
- Constant degree graphs with constant expansion are sparse graphs with good connectivity property.

Theorem

Let $G = (V, E)$ be a regular graph of expansion ϕ. Then, after an $\varepsilon < \phi$ fraction of edges are adversarially removed, the graph has a connected component that spans at least $(1 - \frac{\varepsilon}{2\phi})$ fraction of vertices.

Theorem

Let $G = (V, E)$ be a regular graph of expansion ϕ. Then, after an $\varepsilon < \phi$ fraction of edges are adversarially removed, the graph has a connected component that spans at least $(1 - \frac{\varepsilon}{2\phi})$ fraction of vertices.

- In a d-regular expander (i.e., constant expansion), the removal of k edges can cause $O(k/d)$ vertices to be disconnected from the remaining giant component.
- It is always possible to disconnect k/d vertices after removing k edges. So, the reliability of an expander is the best possible.

End