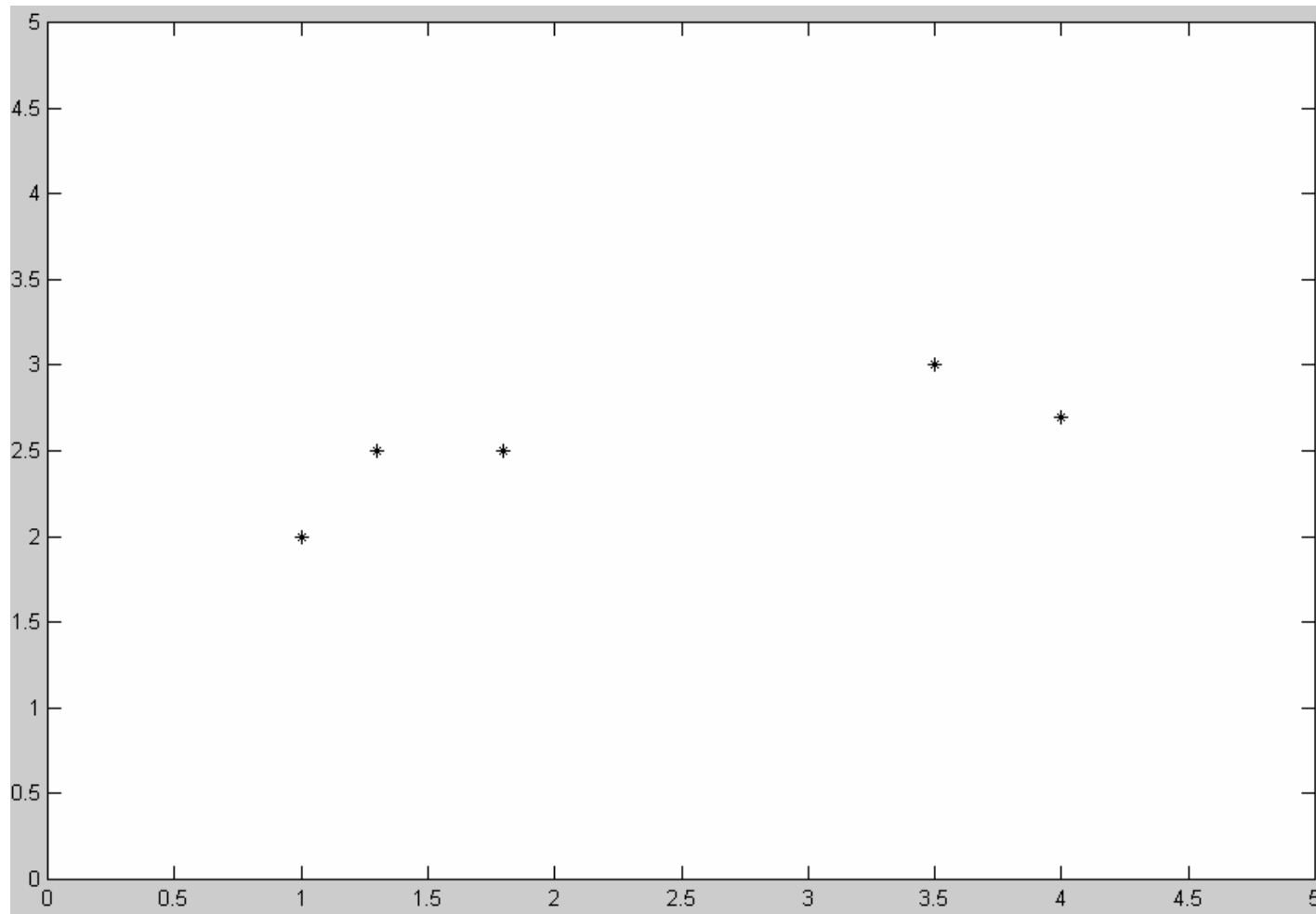


HW3 recitation: Spectral Clustering

Maxim Likhachev

Spectral Clustering

Problem formulation: given n input points, find k clusters



Spectral Clustering

Step 1: Build affinity matrix W

One way to do this:

$$W(i, j) = \begin{cases} 0, & \text{if } x_i \text{ and } x_j \text{ are known to be different} \\ e^{-\beta \cdot \|x_i - x_j\|}, & \text{otherwise} \end{cases}$$

In our example:

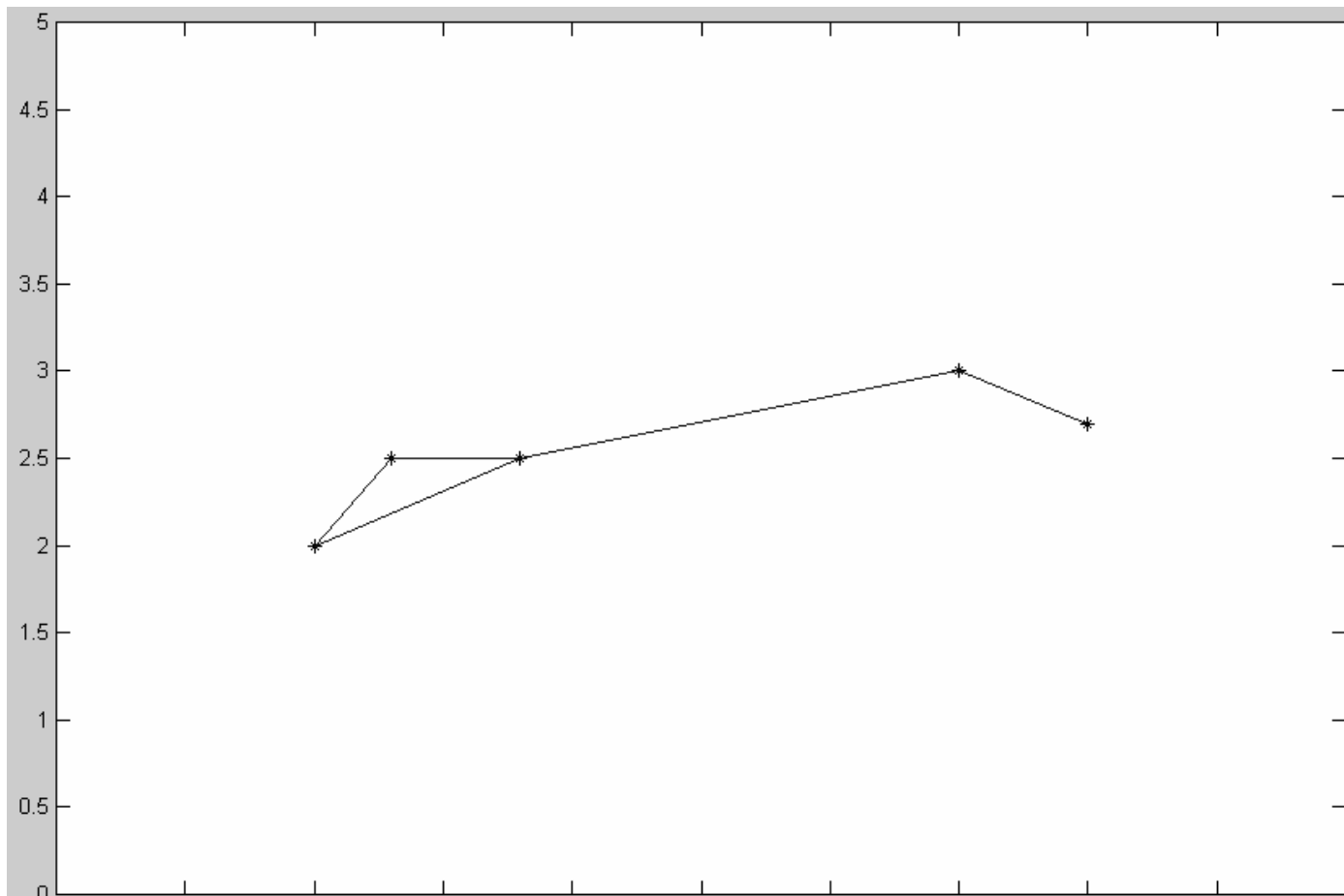
$W =$

| | | | | |
|--------|--------|--------|--------|--------|
| 1.0000 | 0.5582 | 0.3893 | 0 | 0 |
| 0.5582 | 1.0000 | 0.6065 | 0 | 0 |
| 0.3893 | 0.6065 | 1.0000 | 0.1700 | 0 |
| 0 | 0 | 0.1700 | 1.0000 | 0.5582 |
| 0 | 0 | 0 | 0.5582 | 1.0000 |

Spectral Clustering

Step 1: Build affinity matrix W

In our example:



Spectral Clustering

Step 2: Construct $A = D^{-1/2}WD^{-1/2}$, where D – a diagonal matrix with $D_{ii} = \text{sum along } i^{\text{th}} \text{ row in } W$

In our example:

$A =$

| | | | | |
|--------|--------|--------|--------|--------|
| 0.5135 | 0.2719 | 0.1896 | 0 | 0 |
| 0.2719 | 0.4620 | 0.2801 | 0 | 0 |
| 0.1896 | 0.2801 | 0.4617 | 0.0879 | 0 |
| 0 | 0 | 0.0879 | 0.5787 | 0.3401 |
| 0 | 0 | 0 | 0.3401 | 0.6418 |

Spectral Clustering

Step 3: Get eigenvectors of A sorted by their corresponding eigenvalues

In our example:

L (diagonal is eigenvalues) =

| | | | | |
|--------|--------|--------|--------|--------|
| 1.0000 | 0 | 0 | 0 | 0 |
| 0 | 0.9319 | 0 | 0 | 0 |
| 0 | 0 | 0.3297 | 0 | 0 |
| 0 | 0 | 0 | 0.2470 | 0 |
| 0 | 0 | 0 | 0 | 0.1489 |

V =

| | | | | |
|--------|---------|---------|---------|---------|
| 0.4512 | -0.3652 | 0.5757 | -0.5238 | 0.2392 |
| 0.4757 | -0.3730 | 0.0005 | 0.3418 | -0.7195 |
| 0.4759 | -0.2714 | -0.5589 | 0.2460 | 0.5718 |
| 0.4251 | 0.5248 | -0.4035 | -0.5610 | -0.2578 |
| 0.4036 | 0.6152 | 0.4397 | 0.4834 | 0.1779 |

Spectral Clustering

Step 4: Check the signs of the elements in the second eigenvector to split into two clusters

Elements with opposite signs belong to different clusters

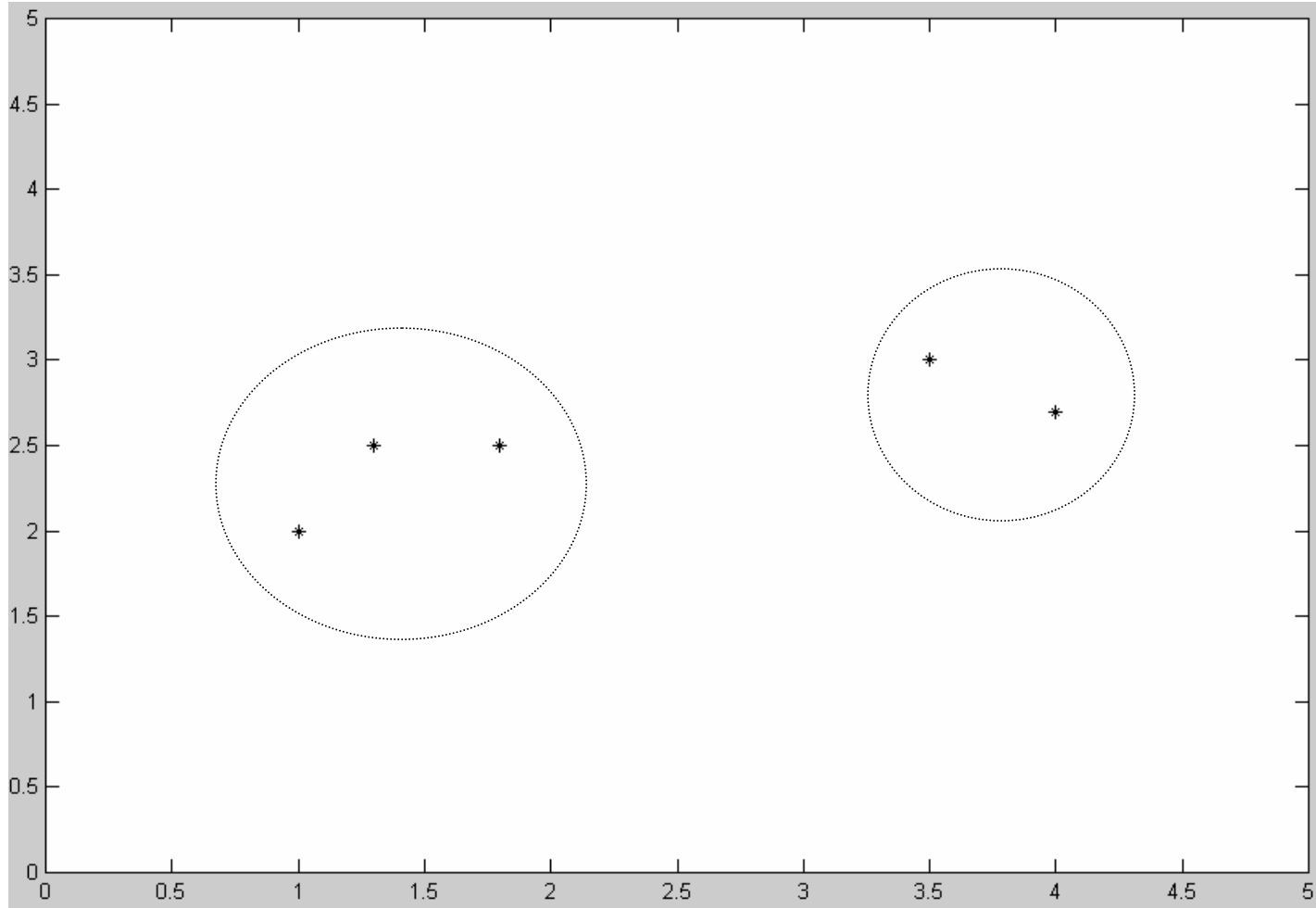
In our example:

$V =$

| | | | | |
|--------|----------------|---------|---------|---------|
| 0.4512 | -0.3652 | 0.5757 | -0.5238 | 0.2392 |
| 0.4757 | -0.3730 | 0.0005 | 0.3418 | -0.7195 |
| 0.4759 | -0.2714 | -0.5589 | 0.2460 | 0.5718 |
| 0.4251 | 0.5248 | -0.4035 | -0.5610 | -0.2578 |
| 0.4036 | 0.6152 | 0.4397 | 0.4834 | 0.1779 |

First three points in cluster 1; second two in cluster 2

Spectral Clustering



Spectral Clustering

Step 1: Build affinity matrix W

Step 2: Construct $A = D^{-1/2}WD^{-1/2}$,
where D – a diagonal matrix with D_{ii} = sum along i th row in W

Step 3: Get eigenvectors of A sorted by their corresponding eigenvalues

Step 4: Elements with opposite signs in the 2nd eigenvector
belong to different clusters

Spectral Clustering

symmetric

Step 1: Build affinity matrix W

Step 2: Construct $A = D^{-1/2}WD^{-1/2}$,
where D – a diagonal matrix with D_{ii} = sum along i th row in W


Step 3: Get eigenvectors of A sorted by their corresponding eigenvalues

Step 4: Elements with opposite signs in the 2nd eigenvector
belong to different clusters

Spectral Clustering

Step 1: Build affinity matrix W


symmetric



Step 2: Construct $A = D^{-1/2}WD^{-1/2}$,

where D – a diagonal matrix with D_{ii} = sum along i th row in W

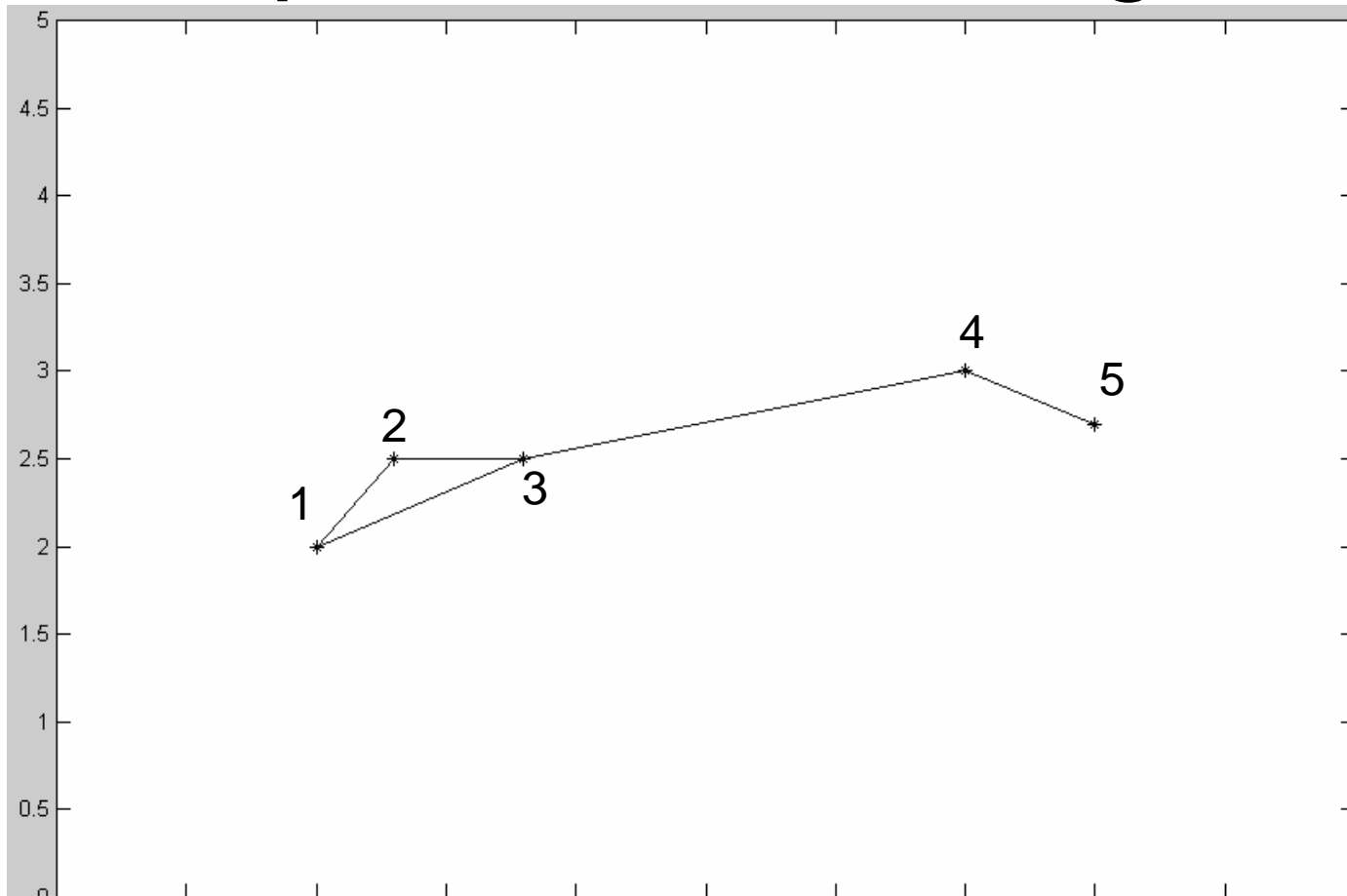
symmetric



Step 3: Get eigenvectors of A sorted by their corresponding eigenvalues

Step 4: Elements with opposite signs in the 2nd eigenvector belong to different clusters

Spectral Clustering



Consider $P = D^{-1}W$: $P_{ij} = w_{ij}/\sum_k w_{ik}$, which is “kind of” the probability of transitioning to point j from point i . (In other words, we normalize each row of P , so that it sums up to 1 and therefore is a valid probability transition matrix.)

Spectral Clustering

In our example:

$P =$

| | | | | |
|--------|--------|--------|--------|--------|
| 0.5135 | 0.2866 | 0.1999 | 0 | 0 |
| 0.2579 | 0.4620 | 0.2802 | 0 | 0 |
| 0.1797 | 0.2800 | 0.4617 | 0.0785 | 0 |
| 0 | 0 | 0.0984 | 0.5787 | 0.3230 |
| 0 | 0 | 0 | 0.3582 | 0.6418 |

$P(\text{end}=2|\text{start}=1 \text{ after 1 step}) = 0.2866;$

$P(\text{end}=i|\text{start}=1 \text{ after 1 step}) = [1 \ 0 \ 0 \ 0 \ 0] * P =$

$[0.5135 \ 0.2866 \ 0.1999 \ 0 \ 0];$

$P(\text{end}=i|\text{start}=1 \text{ after 2 steps}) = [1 \ 0 \ 0 \ 0 \ 0] * P * P =$

$[0.3278 \ 0.3391 \ 0.2973 \ 0.0307 \ 0.0051];$

...

$P(\text{end}=i|\text{start}=1 \text{ after } t \text{ steps}) = [1 \ 0 \ 0 \ 0 \ 0] * P^t;$

Spectral Clustering

$$P^t = (D^{-1}W)^t = D^{-1/2}(D^{-1/2}WD^{-1/2})^t D^{1/2} = D^{-1/2}A^t D^{1/2}$$

$$P^t = D^{-1/2}(\lambda_1^t v_1 v_1^T + \lambda_2^t v_2 v_2^T + \dots) D^{1/2}$$

Since $\lambda_1 = 1$ and $|\lambda_i| < 1$ for $i > 1$,

$$P^\infty = D^{-1/2}(\lambda_1 v_1 v_1^T) D^{1/2}$$

$P_{ij}^\infty = P(\text{end}=j | \text{start}=i \text{ after infinitely many steps})$

Spectral Clustering

$$P^t = (D^{-1}W)^t = D^{-1/2}(D^{-1/2}WD^{-1/2})^t D^{1/2} = D^{-1/2}A^t D^{1/2}$$

$$P^t = D^{-1/2}(\lambda_1^t v_1 v_1^T + \lambda_2^t v_2 v_2^T + \dots) D^{1/2}$$

Since $\lambda_1 = 1$ and $|\lambda_i| < 1$ for $i > 1$,

$$P^\infty = D^{-1/2}(\lambda_1 v_1 v_1^T) D^{1/2}$$

$P_{ij}^\infty = P(\text{end}=j | \text{start}=i \text{ after infinitely many steps})$

How to correct for large number but not infinitely many steps:

$$P^t \approx P^\infty + D^{-1/2}(\lambda_2^t v_2 v_2^T) D^{1/2}$$

In v_2 , $v_2(i)$ has a different sign with $v_2(j)$ if

$P(\text{end}=j | \text{start}=i \text{ after a finite number of steps})$ should decrease in comparison to $P(\text{end}=j | \text{start}=i \text{ after infinitely many steps})$