Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

The Lecture Contains:

- Distances
- Lp norm
- Normalized Euclidean distance
- Quadratic form distance
- Mahalanobis distance
- Kullback-Leibler divergence
- Jensen-Shannon divergence
- Bhattacharyya coefficient

Prev topic Next topic Prev page
Next page

Distances

- Two objects $x = \{x_1, \ldots, x_k\}$ and $y = \{y_1, \ldots, y_k\}$ of dimensionality k
- In general, database of N objects of dimensionality k is represented as a matrix $\mathbf{D}$ of size k x N
- Column-oriented representation

Prev topic Next topic Prev page
Next page

Lp norm

$$L_p(x, y) = \left[\sum_{i=1}^{k} (|x_i - y_i|)^p \right]^{1/p}$$

- Also known as Minkowski norm or Minkowski distance
- L_2:

Prev topic Next topic Prev page
Next page

Lp norm

$$L_p(x, y) = \left[\sum_{i=1}^{k} (|x_i - y_i|)^p \right]^{1/p}$$

- Also known as Minkowski norm or Minkowski distance
- L_2: Euclidean
- L_1:

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Lp norm

$$L_p(x,y) = \left[\sum_{i=1}^{k}(|x_i - y_i|)^p\right]^{1/p}$$

- Also known as Minkowski norm or Minkowski distance
- L_2: Euclidean
- L_1 : Manhattan
- L_∞ :

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Lp norm

$$L_p(x,y) = \left[\sum_{i=1}^{k}(|x_i - y_i|)^p\right]^{1/p}$$

- Also known as Minkowski norm or Minkowski distance
- L_2: Euclidean
- L_1 : Manhattan
- L_∞ : max
- $L_{-\infty}$:

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Lp norm

$$L_p(x,y) = \left[\sum_{i=1}^{k}(|x_i - y_i|)^p\right]^{1/p}$$

- Also known as Minkowski norm or Minkowski distance
- L_2: Euclidean
- L_1 : Manhattan
- L_∞ : max
- $L_{-\infty}$: min
- Metric? Yes for $1 \leq p < \infty$
- Are weighted versions metric distances?

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Lp norm

$$L_p(x,y) = \left[\sum_{i=1}^{k}(|x_i - y_i|)^p\right]^{1/p}$$

- Also known as Minkowski norm or Minkowski distance
- L_2 :Euclidean
- L_1 : Manhattan
- L_∞ : max
- $L_{-\infty}$: min
- Metric? Yes for $1 \leq p < \infty$
- Are weighted versions metric distances? Yes

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Normalized Euclidean distance

$$NED(x,y) = \left[\sum_{i=1}^{k}(x'_i - y'_i)^2\right]^{1/2}$$

- Each dimension is mean-centered and normalized

$$x'_i = (x_i - \mu_i)/\sigma_i$$

- μ_i and σ_i are the mean and standard deviation of dimension i for all data, i.e., the i^{th} row of $\mathbf{D}$
- Metric?

Normalized Euclidean distance

$$NED(x,y) = \left[\sum_{i=1}^{k} (x'_i - y'_i)^2 \right]^{1/2}$$

- Each dimension is mean-centered and normalized

$$x'_i = (x_i - \mu_i)/\sigma_i$$

- μ_i and σ_i are the mean and standard deviation of dimension i for all data, i.e., the i^{th} row of $\mathbf{D}$
- Metric? Yes

Prev topic Next topic Prev page
Next page

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x-y)^T A (x-y)}$$

- A is a $k \times k$ matrix
- A_{ij} denotes the similarity of dimension i with dimension j
- A is positive semi-definite (for distances to be ≥ 0)

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x-y)^T A (x-y)}$$

- A is a $k \times k$ matrix
- A_{ij} denotes the similarity of dimension i with dimension j
- A is positive semi-definite (for distances to be ≥ 0)

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x-y)^T A(x-y)}$$

- A is a $k \times k$ matrix
- A_{ij} denotes the similarity of dimension i with dimension j
- A is positive semi-definite (for distances to be ≥ 0)
  - $z^T Az \geq 0$ for all non-zero vectors z
- A is symmetric (for distances to be symmetric)
- Example: $A_{ij} = 1 - c_{ij}/c_{max}$ for color histograms (c_{ij} is bin-to-bin distance and c_{max} the maximum)
- If A is an identity matrix,

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x-y)^T A (x-y)}$$

- A is a $k \times k$ matrix
- A_{ij} j denotes the similarity of dimension i with dimension j
- A is positive semi-definite (for distances to be ≥ 0)
  - $z^T A z \geq 0$ for all non-zero vectors z
- A is symmetric (for distances to be symmetric)
- Example: $A_{ij} = 1 - c_{ij}/c_{max}$ for color histograms (c_{ij} is bin-to-bin distance and c_{max} the maximum)
- If A is an identity matrix, then Euclidean
- If A is a diagonal matrix,

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Quadratic form distance

$$d_Q(x,y) = \sqrt{(x-y)^T A(x-y)}$$

- A is a $k \times k$ matrix
- A_{ij} j denotes the similarity of dimension i with dimension j
- A is positive semi-definite (for distances to be ≥ 0)
  - $z^T Az \geq 0$ for all non-zero vectors z
- A is symmetric (for distances to be symmetric)
- Example: $A_{ij} = 1 - c_{ij}/c_{max}$ for color histograms (c_{ij} is bin-to-bin distance and c_{max} the maximum)
- If A is an identity matrix, then Euclidean
- If A is a diagonal matrix, then weighted Euclidean
- Metric?

Prev topic Next topic Prev page
Next page

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x-y)^T A(x-y)}$$

- A is a $k \times k$ matrix
- A_{ij} j denotes the similarity of dimension i with dimension j
- A is positive semi-definite (for distances to be ≥ 0)
  - $z^T Az \geq 0$ for all non-zero vectors z
- A is symmetric (for distances to be symmetric)
- Example: $A_{ij} = 1 - c_{ij}/c_{max}$ for color histograms (c_{ij} is bin-to-bin distance and c_{max} the maximum)
- If A is an identity matrix, then Euclidean
- If A is a diagonal matrix, then weighted Euclidean
- Metric? Yes, if A is positive definite

Prev topic Next topic Prev page
Next page

Mahalanobis distance

Replace A in quadratic form by inverse of covariance matrix Σ

$$d_M(x,y) = \sqrt{(x-y)^T \Sigma^{-1} (x-y)}$$

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Mahalanobis distance

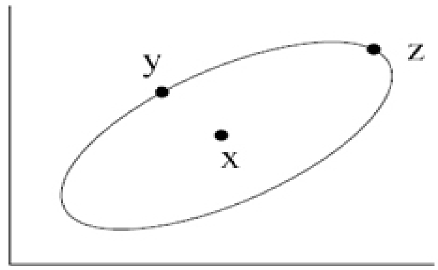
Replace A in quadratic form by inverse of covariance matrix Σ

$$d_M(x,y) = \sqrt{(x-y)^T \Sigma^{-1} (x-y)}$$

- $\sigma_{ij} = E[(x_{\cdot i} - \mu_i)(x_{\cdot j} - \mu_j)] = \sum_{k=1}^{n} \frac{(x_{ki}-\mu_i)(x_{kj}-\mu_j)}{n}$

- Defines an ellipsoid stretched according to strength of each dimension

  - $d_M(x,y) = d_M(x,z)$



- If Σ^{-1} is a diagonal matrix,

Prev topic Next topic Prev page
Next page

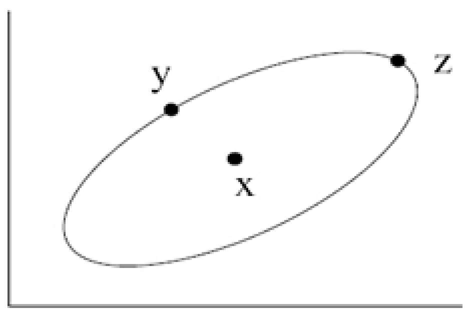
Mahalanobis distance

Replace A in quadratic form by inverse of covariance matrix Σ

$$d_M(x,y) = \sqrt{(x-y)^T \Sigma^{-1}(x-y)}$$

- $\sigma_{ij} = E[(x_{.i} - \mu_i)(x_{.j} - \mu_j)] = \sum_{k=1}^{n} \frac{(x_{ki}-\mu_i)(x_{kj}-\mu_j)}{n}$

- Defines an ellipsoid stretched according to strength of each dimension
  - $d_M(x,y) = d_M(x,z)$



- If Σ^{-1} is a diagonal matrix, then normalized Euclidean
- Metric?

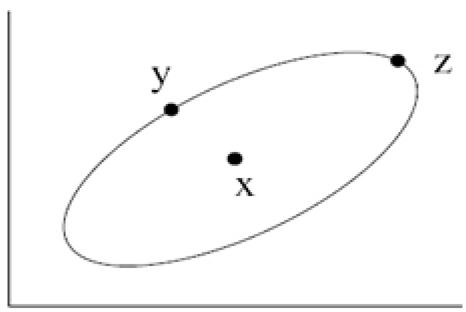
Mahalanobis distance

Replace A in quadratic form by inverse of covariance matrix Σ

$$d_M(x,y) = \sqrt{(x-y)^T \Sigma^{-1} (x-y)}$$

- $\sigma_{ij} = E[(x_{.i} - \mu_i)(x_{.j} - \mu_j)] = \sum_{k=1}^{n} \frac{(x_{ki}-\mu_i)(x_{kj}-\mu_j)}{n}$

- Defines an ellipsoid stretched according to strength of each dimension
  - $d_M(x,y) = d_M(x,z)$



- If Σ^{-1} is a diagonal matrix, then normalized Euclidean
- Metric? Yes

Prev topic Next topic Prev page
Next page

Kullback-Leibler divergence

Also known as Relative entropy or Information divergence

$$d_{KL}(x, y) = \sum_{i=1}^{k} (x_i \log(x_i / y_i))$$

Inefficiency (or difference in bits) when trying to code x using y rather than x itself

- Metric?

Module 2: Distances

Lecture 6: Distance Functions

Kullback-Leibler divergence

Also known as Relative entropy or Information divergence

$$d_{KL}(x, y) = \sum_{i=1}^{k} (x_i \log(x_i/y_i))$$

Inefficiency (or difference in bits) when trying to code x using y rather than x itself

- Metric? No
- Not even symmetric
- Unbounded
    - Bounded for constrained distributions

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Jensen-Shannon divergence

- Also known as Information radius

$$d_{JS}(x,y) = (d_{KL}(x,z) + d_{KL}(y,z))/2$$

- $z = (x+y)/2$
- Attempt to symmetrize KL divergence
- Average KL divergence to average
- Metric?

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Jensen-Shannon divergence

- Also known as Information radius

$$d_{JS}(x,y) = (d_{KL}(x,z) + d_{KL}(y,z))/2$$

- $z = (x+y)/2$
- Attempt to symmetrize KL divergence
- Average KL divergence to average
- Metric? No
- Bounded
- Jeffrey divergence: Twice of Jensen-Shannon divergence

Bhattacharyya coefficient

$$B(x, y) = \sum_{i=1}^{k} \sqrt{x_i y_i}$$

- Measures cosine of angle between direction cosine vectors $(\sqrt{x_1}, \ldots, \sqrt{x_k})$ and $(\sqrt{y_1}, \ldots, \sqrt{y_k})$
- Hellinger distance:

$$d_H(x, y) = 1 - B(x, y)$$

- Metric?

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page Next page

Bhattacharyya coefficient

$$B(x,y) = \sum_{i=1}^{k} \sqrt{x_i y_i}$$

- Measures cosine of angle between direction cosine vectors $(\sqrt{x_1}, \ldots, \sqrt{x_k})$ and $(\sqrt{y_1}, \ldots, \sqrt{y_k})$
- Hellinger distance:

$$d_H(x,y) = 1 - B(x,y)$$

- Metric? Yes

Bhattacharyya coefficient

$$B(x,y) = \sum_{i=1}^{k} \sqrt{x_i y_i}$$

- Measures cosine of angle between direction cosine vectors $(\sqrt{x_1}, \ldots, \sqrt{x_k})$ and $(\sqrt{y_1}, \ldots, \sqrt{y_k})$
- Hellinger distance:

$$d_H(x,y) = 1 - B(x,y)$$

- Metric? Yes
- Bhattacharyya distance:

$$d_B(x,y) = -ln\, B(x,y)$$

- Metric?

Module 2: Distances

Lecture 6: Distance Functions

Prev topic Next topic Prev page
Next page

Bhattacharyya coefficient

$$B(x,y)=\sum_{i=1}^{k}\sqrt{x_i y_i}$$

- Measures cosine of angle between direction cosine vectors $(\sqrt{x_1},...,\sqrt{x_k})$ and $(\sqrt{y_1},...,\sqrt{y_k})$
- Hellinger distance:

$$d_H(x,y)=1-B(x,y)$$

- Metric? Yes
- Bhattacharyya distance:

$$d_B(x,y)=-ln\,B(x,y)$$

- Metric? No