Module 2: Distances

Lecture 8: Earth Mover's Distance (EMD)

Prev topic Next topic Prev page
Next page

The Lecture Contains:

- General case
- EMD general formulation
- Match distance
- Edit distance

General case

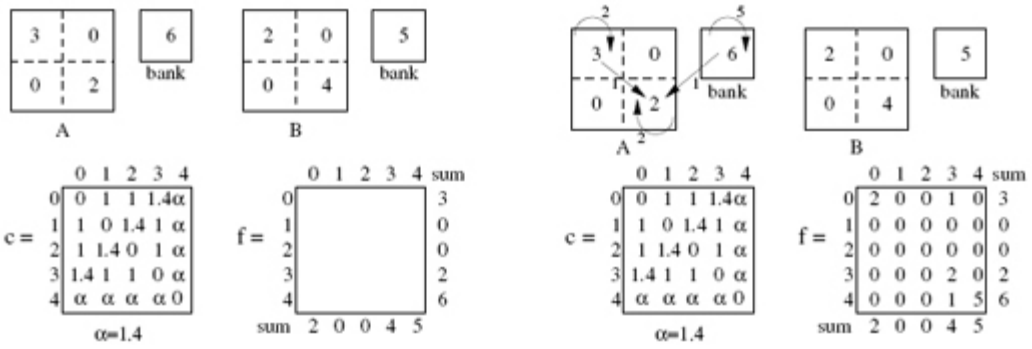
- Feature values need to be positive
  - Done by translating the entire feature range
- Feature values in x and y need to add up to the same mass
  - Normalization may not always be suitable

General case

- Feature values need to be positive
  - Done by translating the entire feature range
- Feature values in x and y need to add up to the same mass
  - Normalization may not always be suitable
- Special "bank" region
- Bank of x is sum of feature values of y and vice versa
- Parameter α: ground distance from ordinary regions to bank
- Bank to bank distance is 0
- $\mathbf{C}$ and $\mathbf{F}$ become matrices of size $(m+1) \times (n+1)$
- Metric?

General case

- Feature values need to be positive
  - Done by translating the entire feature range
- Feature values in x and y need to add up to the same mass
  - Normalization may not always be suitable
- Special "bank" region
- Bank of x is sum of feature values of y and vice versa
- Parameter α: ground distance from ordinary regions to bank
- Bank to bank distance is 0
- $\mathbf{C}$ and $\mathbf{F}$ become matrices of size $(m+1) \times (n+1)$
- Metric? Yes, if ground distance is a metric



Module 2: Distances

Lecture 8: Earth Mover's Distance (EMD)

Prev topic Next topic Prev page
Next page

EMD general formulation

$$EMD(x, y, \alpha) = \min_{F} \sum_{i=1}^{m+1} \sum_{j=1}^{n+1} (c_{ij} f_{ij})$$

$$\forall i, j, f_{ij} \geq 0;\ \forall i, \sum_{j=1}^{n+1} f_{ij} = x_i;\ \forall j, \sum_{i=1}^{m+1} f_{ij} = y_j$$

Match distance

- A special form of earth mover's distance
- Only one dimension in ground distance
- Between histograms or vectors adding up to same total mass
- Ground distance between bins i and j is $c_{ij} = |i - j|$

Module 2: Distances

Lecture 8: Earth Mover's Distance (EMD)

Prev topic Next topic Prev page
Next page

Match distance

- A special form of earth mover's distance
- Only one dimension in ground distance
- Between histograms or vectors adding up to same total mass
- Ground distance between bins i and j is $c_{ij} = |i - j|$

$$d_M(x, y) = L_1(X, Y)$$

- X and Y denote cumulative distributions
- Metric?

Module 2: Distances

Lecture 8: Earth Mover's Distance (EMD)

Prev topic Next topic Prev page
Next page

Match distance

- A special form of earth mover's distance
- Only one dimension in ground distance
- Between histograms or vectors adding up to same total mass
- Ground distance between bins i and j is $c_{ij} = |i - j|$

$$d_M(x, y) = L_1(X, Y)$$

- X and Y denote cumulative distributions
- Metric? Yes

Edit distance

- Also known as Levenshtein distance
- Used for words and strings
- Minimum number of edit operations required to transform word x to word y
- Edit operations on characters are
  - Insertion
  - Deletion
  - Substitution
- Insertions and deletions can be treated as substitutions using gap
- Example: each edit operation costs 1

idea

deal

Edit distance

- Also known as Levenshtein distance
- Used for words and strings
- Minimum number of edit operations required to transform word x to word y
- Edit operations on characters are
  - Insertion
  - Deletion
  - Substitution
- Insertions and deletions can be treated as substitutions using gap
- Example: each edit operation costs 1

idea idea-

⇒

deal -deal

- Edit distance is 2
- Metric?

Edit distance

- Also known as Levenshtein distance
- Used for words and strings
- Minimum number of edit operations required to transform word x to word y
- Edit operations on characters are
  - Insertion
  - Deletion
  - Substitution
- Insertions and deletions can be treated as substitutions using gap
- Example: each edit operation costs 1

idea idea-

⇒

deal -deal

- Edit distance is 2
- Metric? Yes