Short Notes On Similarity/Dissimilarity Measures

Ng Yen Kaow

Distance/Dissimilarity & Similarity

- Let d_{ij} denote the **distance/dissimilarity** between two objects x_i and x_j
  - The objects are, for example, strings, sequences, structures, words, documents, pixels, or vectors (of features)
- Similarly s_{ij} denotes the **similarity** between x_i and x_j
- Comparing some objects are better done with a similarity measure, while comparing some other objects are better with a dissimilarity measure

Desirable properties

- Conditions for metric distance
  - $d_{ij} \geq 0$ (non-negativity)
  - $d_{ij} = 0$ if and only if $i = j$ (identity of indiscernible pairs)
  - $d_{ij} = d_{ji}$ (symmetry)
  - $d_{ij} \leq d_{ik} + d_{kj}$ (triangular inequality)
- Similar intuitions for other dissimilarity (or similarity) measures
- Many dissimilarity/similarity measures can be defined

Examples of dissimilarity measures

- Strings/Sequences
  - Hamming distance
  - Sequence alignment, e.g. edit distance
- Structure
  - Root Mean Square Deviation (RMSD)
  - See https://en.wikipedia.org/wiki/Structural_alignment

Hamming distance, edit distance, RMSD are all metric

- **Vectors**
  - Euclidean distance
  - Metric distance
  - Non-metric distance

Examples of similarity measures

- Words/Documents
  - Bag-of-words, tf-idf
  - Semantic (https://en.wikipedia.org/wiki/Semantic_similarity)
  - Vector (https://en.wikipedia.org/wiki/Word_embedding)
- **Vectors**
  - Correlations (Pearson *etc.*)
  - Covariance
    - Principal Component Analysis
  - Gaussian $e^{-\|x_i - x_j\|^2/2\sigma^2}$
    - Mapping to infinite dimensional space (Kernel function)
    - Probability distribution (co-occurrence probability)
    - Heat function (transition probability)

Gaussian function

- The Gaussian function is

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2}$$

  - Used prominently in
    - Kernel methods
    - Image segmentation (Wu and Leary 1993, Normalized Cut 1997)
    - Dimensionality reduction (Eigenmap 2003, Diffusion maps 2005, t-SNE 2007, UMAP 2018)
- Pros:
  - Linear combination of $(x_i^\top x_j)^k$ terms for all powers of k
  - Fast decay to zero
  - Symmetric, non-negative, identity
- Con: Sensitive to σ

Converting $d_{ij} \Leftrightarrow s_{ij}$

- Difficult to obtain s_{ij} from d_{ij} and vice versa
  - Most conversions will be dissatisfactory, resulting in non-metric distance
- **Ad hoc conversion** between dissimilarity $D = (d_{ij})$ and similarity $S = (s_{ij})$
  - Inverse conversion
    - $d_{ij} = \text{const} * (1 + s_{ij})^{-1}$
    - $s_{ij} = \text{const} * (1 + d_{ij})^{-1}$
  - Linear conversion
    - $d_{ij} = \text{const} - s_{ij}$
    - $s_{ij} = \text{const} - d_{ij}$

Euclidean distance $d_{ij} \Leftrightarrow s_{ij}$

- Let $D = (d_{ij})$ be given by the Pythagorean

$$d_{ij}^2 = (x_i - x_j)(x_i - x_j)^\top$$

  where x_i and x_j are row vectors

- For $S = (s_{ij})$
  - Cosine similarity

$$s_{ij} = \frac{x_i x_j^\top}{\|x_i\| \|x_j\|}$$

  - Linear kernel similarity

$$s_{ij} = x_i x_j^\top$$

    - Con: $s_{ij} \leq s_{uv}$ does not imply $d_{ij} \geq d_{uv}$
    - Pro: Can be converted to d_{jk} easily (next slide)

Euclidean distance $d_{ij} \Leftrightarrow s_{ij}$

- If s_{ij} is the linear kernel similarity, that is, $s_{ij} = x_i x_j^\top$
  - $d_{ij}^2 = (s_{ii} + s_{jj}) - 2s_{ij}$
  - $S = -\frac{1}{2} CDC$

    where

    $C = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, the centering matrix

    $\mathbf{1}$ is a column vector of all ones (hence $\mathbf{1}\mathbf{1}^\top$ is a matrix with all ones of the same dimension as D)
- No similar relation exists for the cosine distance (use ad hoc)

Gaussian similarity $s_{ij} \Leftrightarrow d_{ij}$

- For Gaussian similarity $S = (s_{ij})$ and dissimilarity $D = (d_{ij})$
  - $s_{ij} = e^{-\frac{d_{ij}^2}{2\sigma^2}}$
  - Intuitively $d_{ij} = -\alpha \log(s_{ij})$
    - $d_{ii} = 0$ and $d_{ij} = d_{ji}$
  - Alternatively, define an induced distance $d'_{ij} = s_{ii} + s_{jj} - 2s_{ij}$, then
    - $d'_{ij} = 2(1 - s_{ij})$
      - $d'_{ii} = 0$
      - $d'_{ij} = d'_{ji}$

      But still no triangular inequality guarantee