Applied Machine Learning

Low Dimensional Embeddings

Low-Dimensional Embeddings

- Reminder of Principal Coordinate Analysis (PCoA)
- Sammon Mapping
- Stochastic Neighbor Embedding

Low-Dimensional Embedding

- $\mathbf{x}_i \mapsto \mathbf{y}_i$
  - N items
  - $\mathbf{x}$: high-dimensional dataset with d features $\qquad \mathbf{x}_i = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_d \end{bmatrix}$
  - $\mathbf{y}$: low-dimensional dataset with m features $\qquad \mathbf{y}_i = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{bmatrix}$
- usually $d \gg m$
  - $m \in \{2,3\}$ for visualization

Principal Coordinate Analysis

- Preserves rations of distances between sets $\mathbf{x}$ and $\mathbf{y}$
  - Distances between items within set $D_{i,j}(\mathbf{x}) = (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$
  - Select items in $\mathbf{y}$ that minimize cost function: $\sum_{i,j} (D_{i,j}(\mathbf{x}) - D_{i,j}(\mathbf{y}))^2$
- Issue: weight of pairs of points vary with their distance
  - Pairs with long distances have a higher squared distance
  - Pairs with small distances have smaller squared distance
  - uneven distribution of distances in lower-dimensional map

Sammon Mapping

- Sammon mapping function gives higher weight to smaller distances
- $$C(\mathbf{y}) = \frac{1}{\sum_{i<j} \|\mathbf{x}_i - \mathbf{x}_j\|} \sum_{i<j} \frac{(\|\mathbf{y}_i - \mathbf{y}_j\| - \|\mathbf{x}_i - \mathbf{x}_j\|)^2}{\|\mathbf{x}_i - \mathbf{x}_j\|}$$
- Issue:
  - critical to correctly map pairs much closer to each other than every other pair
- Solved through gradient descent

Stochastic Neighbor Embedding (SNE)

- Preserve the probability of neighbors in set $\mathbf{x}$ in the target set $\mathbf{y}$
- Probability models:
  - probability of items in high-dimensional set $\mathbf{x}$ to be neighbors
  - probability of items in low-dimensional set $\mathbf{y}$ to be neighbors

SNE - model for high-dimensional set

- high-dimensional set $\mathbf{x}$ with N items
- probability of item i being picked by point j as a neighbor in high-dimensional dataset $\mathbf{x}$

$$p_{j|i} = \frac{e^{-\frac{\|\mathbf{x}_j - \mathbf{x}_i\|^2}{2\sigma_i^2}}}{\sum_{k \neq i} e^{-\frac{\|\mathbf{x}_k - \mathbf{x}_i\|^2}{2\sigma_i^2}}}$$

- variance as length scale for point i: σ_i^2
  - chosen so that $p_{i,j}$ has user-defined perplexity $PP(p) = 2^{H(p)}$ with $H(p)$: entropy of p
  - larger σ_i^2: more neighbors around i
  - smaller σ_i^2: less neighbors around i
- probability of items in high-dimensional dataset $\mathbf{x}$ (with N items) to be neighbors:

$$p_{i,j} = \begin{cases} 0 & i = j \\ \frac{p_{j|i} + p_{i|j}}{2N} & i \neq j \end{cases}$$

SNE - model for low-dimensional set

- low-dimensional set $\mathbf{y}$ with N items
- crowding problem: distances in low dimensions reduce with respect to high dimensions
  - Probabilistic model with heavy tails reduce crowding in low dimensions
  - t-SNE
    - Student's t distribution with $\nu = 1$
      - similar to normal with higher probability far from the mean
- probability of items in low-dimensional dataset $\mathbf{y}$ to be neighbors: $$q_{i,j} = \begin{cases} 0 & i = j \\ \dfrac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}} & i \neq j \end{cases}$$

t-SNE Cost Function and Gradient

- Goal: high-dimensional $p_{i,j}$ similar to low-dimensional $q_{i,j}$
  - KL Divergence between P and Q $$\mathbb{D}\left(P||Q\right) = \int P(X) log \frac{P(X)}{Q(X)} dx$$
  - Cost function to minimize $$\mathscr{L} = C_{tSNE} = \sum_{i \neq j} p_{i,j} \log \frac{p_{i,j}}{q_{i,j}}$$
- Gradient descent
  - Gradient $$\nabla y_i \mathscr{L} = 4 \sum_j \frac{(p_{i,j} - q_{i,j})(\mathbf{y}_i - \mathbf{y}_j)}{1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2}$$

Low-Dimensional Embeddings

- Reminder of Principal Coordinate Analysis (PCoA)
- Sammon Mapping
- Stochastic Neighbor Embedding

Applied Machine Learning

Low Dimensional Embeddings