neighbor retrieval visualizer - NeRV
Projections

linear

- PCA
- ICA
- Projection Pursuit

non-linear

- deterministic
  - Distance-Error
    - MDS
    - Sammon's Mapping
  - graph based
    - ISOMAP
    - LLE

- stochastic
  - focusing
    - CCA
    - GTM
    - t-SNE
  - databionic
    - ESOM
    - ABC
    - DBS

NerV
Grundlagen

- **Input Space** $I \subset \mathbb{R}^n$, $D(l,j)$ Distanz
- **Output Space** $O \subset \mathbb{R}^2$, $d(l,j)$ Distanz
- **Projektion:** $proj: I \rightarrow O$, $D(l,j) \leftrightarrow d(l,j)$, where $a l$ and $j$ are points in the corresponding metric spaces

- Backward Projection Error (BPE)
- Forward Projection Error (FPE)
Problem

Bei einer Projektion $\mathbb{R}^n \rightarrow \mathbb{R}^m$, $m << n$ können NIE alle Nachbarschaften oder gar Distanzen perfekt erhalten werden

$\rightarrow$ ShepardDiagram
Good if:

Points that are close in the output space are close in the original space

http://www.uta.fi/sis/mtt/mtts1-dimensionality_reduction/...
2014Peltonen_drv_lecture7
Recall – Quality for FPE

Good if:

Points that are close in the input space are close in the output space
Prinzip: NeRV

- Lyapunov Funktion E (objective function)
- Idee: SNE mit Precision und Recall und kaum Fokussierung

Zuerst: Nachbahrshaftsdefinition…
“Hard” Neighborhoods

\[
\text{Precision}(i) = 1 - \frac{N_{\text{False positive}}}{k_i}
\]

\[
\text{Recall}(i) = 1 - \frac{N_{\text{Miss}}}{r_i}
\]
Motivation: “Soft” Neighborhoods

- “In vectorial data, if nothing else is known, it is reasonable that close-by points in some metric can be considered neighbors.
- each point is a neighbor with some weight and a non-neighbor with some Weight”

http://www.uta.fi/sis/mtt/mtts1-dimensionality_reduction/... 2015Peltonen_MTTS_lecture10:
Probabilistic neighborhood

- die Wahrscheinlichkeit, dass Datenpunkt j den Punkt j zum Nachbarn hat

\[ p(l,j) = \frac{\exp\left(-\frac{D(l,j)^2}{2\sigma^2}\right)}{\sum_{l \neq j} \exp\left(-\frac{D(l,j)^2}{2\sigma^2}\right)} \]

\[ q(l,j) = \frac{\exp\left(-d(l,j)^2/2\sigma^2\right)}{\sum_{l \neq j} \exp\left(-d(l,j)^2/2\sigma^2\right)} \]

⇒ The same as SNE

(Stochastic Neighbor Embedding)
Two probability distributions over a set of items can be compared by the **Kullback-Leibler (KL) divergence**

- relative entropy
- amount of surprise when encountering items from the 1\textsuperscript{st} distribution when items were expected to come from the 2\textsuperscript{nd}

\[
E_{SNE} = \sum_{l} \sum_{j \neq l} p(l,j) \log \left( \frac{p(l,j)}{q(l,j)} \right) := KL(p)
\]

http://www.uta.fi/sis/mtt/mtts1-dimensionality_reduction/...
SNE - Details

- Minimization of Objective function $E$
- nonnegative, and zero if and only if the distributions are equal
- value of the divergence sum depends on output coordinates, and can be minimized with respect to them
- $\sigma$ as radius if neighborhood
In NeRV

“Scale parameter” $\sigma$ fest

- Peltonen: effective number of neighbors
- Peltonen: $c(j) = \log(k(j))$,

$k$ - retrieved points (green) around point $j$

http://www.uta.fi/sis/mtt/mtts1-dimensionality_reduction/...
2015Peltonen_MTTS_lecture10:
Äquivalenz Recall R

\[ R(l) = 1 - \frac{N_{\text{Miss}}}{r_l} \leftrightarrow \sum_l \sum_{j \neq l} p(l, j) \log \left( \frac{p(l,j)}{q(l,j)} \right) \]

Proof: 2010, Venna et al. Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization

Peltonen: “assume some probabilities are uniformly-large, some are uniformly-small. Then there are 4 different kinds of terms in the sum. Show that above KL divergence is dominated by a cost that is proportional to a constant times number of misses”
"Soft" Neighborhoods II

\[ p(l,j) \log \left( \frac{p(l,j)}{q(l,j)} \right) = p(l,j) \times \text{weight} \]

=> \( R \sim \) sum over observed center-neighbor pairs, weighted by their proportional counts \( p_{ij} \).

We sum the log-likelihoods of those observations (2015Peltonen_MTTS_lecture10)
Folgerung: Precision Pr

- “SNE focuses on recall (misses) because its cost function is dominated by misses”

\[ Pr(l) = 1 - \frac{N_{\text{False positive}}}{k_l} \iff \sum_l \sum_{j \neq l} q(l, j) \log \left( \frac{q(l,j)}{p(l,j)} \right) \]

- “change the retrieval model so that misses become less dominant, so that the model can also focus on false positives”

Proof: 2010, Venna et al. Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization

Peltonen: “assume some probabilities are uniformly-large, some are uniformly-small. Then there are 4 different kinds of terms in the sum. Show that above KL divergence is dominated by a term proportional to a constant times number of false neighbors”
Neighbor Retrieval Visualizer

\[ E_{NeRV} = \lambda \ast Ex\{KL(p)\} + (1 - \lambda) \ast Ex\{KL(q)\} \]

- Minimize with respect to output space points
- \(Ex \sim \text{“Expectation” (Mittelwert)}\)
- \(\lambda \sim \text{trdof between Precision and Recall}\)
- User has to choose \(\lambda\)

1.) Initialization “To speed up convergence and avoid local minima”

2.) standard conjugate gradient step (analog MDS)

2010, Venna et al. Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization
Details NeRV

- 20 steps with decreasing $\sigma$ per two steps
- Linear decreasing
  $\sigma_{max} \sim \text{half diameter of input data}$
  $\sigma_{min} = \log(k(j))$

After Initialization
- 20 standard conjugate gradient steps with $\sigma_{min}$
\[ \lambda \sim \text{Tradoff between Pr and R} \]

- A: \( \lambda \) niedrig -> Precision hoch
  "minimizes false positives (false retrieved neighbors)"

- b: \( \lambda \) hoch -> Recall hoch
  "minimizes misses (neighbors that were not retrieved)"

2010, Venna et al. Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization
"In unsupervised visualization, NeRV outperformed alternatives for most of the six data sets we tried, for four different pairs of measures, and was overall the best method."

- Verglichen wurde mit PCA, MDS, LLE, Laplacian eigenmap, Hessian-based locally linear embedding, isomap, curvilinear component analysis (CCA), curvilinear distance analysis (CDA), maximum variance unfolding (MVU), landmark maximum variance unfolding (LMVU), and local MDS (LMDS), LE, HLLE

- Bewiesen wird mit Trustworthiness und Continuity

„NeRV also performed well in a comparison by unsupervised classification.“ (Letter, Phoneme, Landsat, TIMIT)

NeRV for supervised visualization (distance = Riemannian topology preserving metric) für 4 Datensätze
FCPS und echte Daten

- Problem: $\mathbb{R}^2 \to \mathbb{R}^2$ ändert sich nichts
- Echter Datensatz von 10000 Zeilen:
  - Bad_alloc Fehler (Speicherproblem)
- Echter Datensatz: 800 Dimension in
  D:\Subversion\lehre\Vorlesungen\KnowledgeDiscovery\01Text\04ProjektionenUndVisualisierung\53NichtlineareProjektionenNeRV\Beispiele.R