Manifold Masking

- Emulate criteria used in linear/nonlinear embedding algorithms (Isomap, NuMax) to obtain structure-preserving masking patterns for manifold-modeled data.
- Seek masking index set $\mathcal{O} = \{\omega_1, \ldots, \omega_m\}$ that is a subset of the dimensions $|d| = \{1, 2, \ldots, d\}$ of $\mathbf{x} \subset \mathbb{R}^d$.
- Define masking linear operator $\psi: \mathbf{x} \mapsto (\omega_i)_{i \in \mathcal{O}}$ corresponding to masking index set $\mathcal{O}$.

Manifold Learning and Linear Dimensionality Reduction:

- An $r$-dimensional manifold $\mathcal{M} \subset \mathbb{R}^d$ is a set of data points $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \subset \mathbb{R}^d$ that have been generated according to an $r$-dimensional parametric function.
- Goal: Given high-dimensional data set $\mathcal{X}$, find underlying parameterization of the manifold $\mathcal{M} \subset \mathbb{R}^d$.

- **Dimensionality reduction**: embed data $\mathcal{X}$ to low-dimensional space $\mathbb{R}^r$ ($m \ll d$) so that local geometry of $\mathcal{M}$ is preserved, i.e., distances in $\mathbb{R}^r$ correspond to parameter differences in $\mathbb{R}^d$.
- **Linear dimensionality reduction**: use a matrix projection $\Phi \in \mathbb{R}^{d \times r}$, e.g., principal component analysis (PCA), multidimensional scaling (MDS).
- PCA/MDS fail to preserve geometric structure of a nonlinear manifold.

Optimization-Based Mask Selection:

- Minimize distortion incurred by secants with neighboring $k$ data points:
  $$S_k = \left\{ \frac{\mathbf{x}_i - \mathbf{x}_j}{|\mathbf{x}_i - \mathbf{x}_j|_2} : i \in [d], j \in \mathcal{N}_k(\mathbf{x}_i) \right\} \subseteq \mathcal{S}.$$  
  - Expectation of masked secant norms over masking index sets $\mathcal{O}$ drawn uniformly at random is $\mathbb{E} [\| \mathbf{w}_{\mathcal{O}} \|_2] = \frac{d}{2}$.

- Secants $S_k$ inevitably subject to compaction factor of $\sqrt{\frac{d}{2}}$ in expectation by masking operator $\psi$.
- Seek masking $\psi$ such that for all $\mathbf{a} \in S_k$, the squared norm of masked secants $\| \mathbf{w}_{\mathcal{O}} \|_2$ is as close as possible.
- We have $\| \mathbf{w}_{\mathcal{O}} \|_2 = \sum_{j \in \mathcal{O}} \mathbf{a}_j \mathbf{z}_j = \sum_{j \in \mathcal{O}} \mathbf{a}_j (2 - z_j^2) - \mathbf{a}_0^2 z_j^2$, where $\mathbf{a}_0^2 = \mathbf{a}_j$ entrywise and $z$ is the $d$-dimensional indicator vector for index set $\mathcal{O}$, i.e., $z_\mathcal{O} = 1$ if $j \in \mathcal{O}$, $0$ otherwise.
- Squared secants matrix $\mathbf{A} = \{S_k \times d\}$ matrix defined by
  $$\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \cdots] = [\mathbf{a}_1 \mathbf{a}_2^2 \cdots]^T.$$  
- Find optimal masking pattern by casting the following integer program:
  $$z^* = \arg \min \mathbb{E} [\| \mathbf{w}_{\mathcal{O}} \|_2] = \mathbb{E} [\| \mathbf{w}_{\mathcal{O}} \|_2] = \mathbb{E} [\| \mathbf{w}_{\mathcal{O}} \|_2],$$  
  subject to $z^* \leq m$, $z \in \{0, 1\}^d$, where $\mathbf{1}_m$ denotes $d$-dimensional all-ones vector.
- Equality constraint $\mathbf{1}_m z = m$: only $m$ dimensions are to be retained in the masking process.
- Integer program (1) is computationally intractable (run only for 24 hours in experiments).

Manifold-Aware Pixel Selection (MAPS)

- **Inputs**: normalized squared secants matrix $\mathbf{A}$, number of dimensions $m$.
- **Outputs**: masking index set $\mathcal{O}$.

**Initialize**: $\mathcal{O} = \{\}$

- **for** $i = 1$ to $m$ **do**
  - $\mathbf{A}_i \leftarrow \mathbf{A}_i + \mathbf{1}_{[1]}$ (compute current masked secant square norms)
  - $\omega_i \leftarrow \arg \min_{\omega_i} \left\| \mathbf{A}_i + \mathbf{A}_i - \mathbf{A}_i \parallel \omega_i \parallel_2 \right\|_2$ (minimize aggregate difference with $\mathbb{E} [\| \mathbf{w}_{\mathcal{O}} \|_2]$)
  - $\mathcal{O} \leftarrow \mathcal{O} \cup \omega_i$ (add selected dimension to the masking index set)

- Heuristic greedy algorithm that can find an approximate solution for (1) in drastically reduced time (seconds/minutes).
- MAPS iteratively selects elements of the masking index set $\mathcal{O}$ using the squared secants matrix $\mathbf{A}$. At iteration $i$ of the algorithm, MAPS finds a new dimension that, when added to the existing dimensions in $\mathcal{O}$, causes the squared norm of the masked secants to match the expected value of $\frac{d}{2}$ as closely as possible on average.
- Computational complexity of MAPS is $O(m^d |S_k|) = O(mkn)$.

Simulation Results

- Compare (1) and MAPS with two baseline methods: random masking (select $m$ out of $d$ data dimensions uniformly at random) and principal coordinate analysis (PCoA) (select indices of $m$ dimensions with the highest variance across the dataset).
- Eyeglasses dataset: eye-tracking image captures via computational eyeglasses prototype that uses pixel-level imaging sensor array of [3].
- MAPS significantly outperforms random sampling and PCoA.
- For sufficiently large values of $m$, the performance of MAPS approaches or matches that of nonlinear linear embedding algorithms on full images.

Acknowledgments and References

This work was supported by NSF Grant IIS-1239341.