Interactive Dimensionality Reduction with User Feedback

Marco Cusumano-Towner
General Idea

- The user positions some data points on a plane.
- The organization of these points (e.g. pairwise distances) gives information about what structure the user is interested in.
- Algorithm learns a distance metric that tries to encodes this structure.
- Algorithm re-projects remaining points using new distance metric.
- Repeat.
Contrived Example

Original data

User is interested in separating data points based on x1
User-defined control points
User pulls apart control points
Algorithm learns metric that weights $x_1$ greater than $x_2$. 
Global effect
Relation to Prior Work

- Prior work re-projects the whole dataset, including the control points.
  - Ideally, we want smooth animations, so user understands the dynamics. Also, can we fix the control points?
- Prior work is limited in the family of distance metrics that can be learned (simply weighting of original dimensions).
  - More flexible metric learning scheme?
Current Algorithm

In realtime:
1. User may move control points
2. One gradient descent step of MDS, with control points fixed
3. One gradient descent step of metric learning, using current control points
4. Repeat

- Results in a **fluid interface**, without jarring changes. Continuous back-and-forth feedback
Next Steps

Tested on Iris dataset, but need more complex metric learning to get good separability. Instead, will focus on data where the *individual dimensions are more meaningful*.

1. Text Document Data (Abstracts)
   - Original dimensions are words or stems (or more complex features)
   - Learning metric weights corresponds to learning which words the user cares about.

2. Gene Annotations
   - Original dimensions are functional annotation terms
Questions for Feedback

• Any ideas for other types of data that would be suited to this type of data exploration method?
• (Data where it would be useful for user to pick and poke at data to test their hypotheses)