Fitting a Graph to Vector Data

Samuel I. Daitch (Yale)
Jonathan A. Kelner (MIT)
Daniel A. Spielman (Yale)

Machine Learning Summer School, June 2009
Given a collection of vectors \( x_1, \ldots, x_n \in \mathbb{R}^d \)

Find a natural weighted graph on \( V = \{1, \ldots, n\} \) that is

1. Theoretically well-motivated
2. Helps solve ML problems on the vectors
3. Has interesting combinatorial properties
4. Efficiently computable
5. Sparse
Given a collection of vectors \( x_1, \ldots, x_n \in \mathbb{R}^d \)

Find a natural weighted graph on \( V = \{1, \ldots, n\} \) that is

1. Theoretically well-motivated
2. Helps solve ML problems on the vectors
2'. Helps solve TCS problems on the vectors
3. Has interesting combinatorial properties
4. Efficiently computable
5. Sparse
Outline

Standard graphs
Motivation for ours
Laplacian view
Combinatorial properties (Sparsity & Planarity)
Use in classification, regression and clustering
Efficient computation

Approximate sparsity
Lovasz’s graphs

Open Questions
Standard ways to make graphs

Choose one from each column

<table>
<thead>
<tr>
<th>Choice of edges</th>
<th>Weights of edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>k nearest neighbors</td>
<td>unweighted</td>
</tr>
<tr>
<td>neighbors if distance less than</td>
<td>$\exp \left( - \frac{</td>
</tr>
<tr>
<td>threshold $\delta$</td>
<td></td>
</tr>
</tbody>
</table>
Our first proposal

Choose weights to minimize

\[ \sum_{i=1}^{n} \left\| d_i \mathbf{x}_i - \sum_{j \sim i} w_{i,j} \mathbf{x}_j \right\|^2 \]

where the weighted degrees satisfy

\[ d_i = \sum_{j \sim i} w_{i,j} \geq 1 \]
Our first proposal

Choose weights to minimize

$$\sum_{i=1}^{n} \left\| d_i \mathbf{x}_i - \sum_{j \sim i} w_{i,j} \mathbf{x}_j \right\|^2$$

$$w_{i,j} = w_{j,i}$$

where the weighted degrees satisfy

$$d_i = \sum_{j \sim i} w_{i,j} \geq 1$$
Motivation

Consider regression problem \( y_i = f(x_i) \) solving for \( y_i \) given \( \{y_j : j \neq i\} \)

Given a weighted graph, natural guess is

\[
\frac{1}{d_i} \sum_{j \sim i} w_{i,j} y_j
\]

Sum of squares of errors when leave out each \( i \) is

\[
\sum_{i=1}^{n} \left( y_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} y_j \right)^2
\]
Motivation

Try to get right on coordinate vectors,

\[ \mathbf{x}_i = (x_{i1}, \ldots, x_{id}) \]

Sum of squares of errors over \( d \) coordinate vectors

\[
\sum_{k=1}^{d} \sum_{i=1}^{n} \left( x_{ki} - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x_{kj} \right)^2
\]

\[
= \sum_{i=1}^{n} \left\| \mathbf{x}_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} \mathbf{x}_j \right\|^2
\]
Motivation

Try to get right on coordinate vectors,

\[ \mathbf{x}_i = (x_i^1, \ldots, x_i^d) \]

Sum of squares of errors over d coordinate vectors

\[
\sum_{k=1}^{d} \sum_{i=1}^{n} \left( x_i^k - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x_j^k \right)^2
\]

\[
= \sum_{i=1}^{n} \left\| \mathbf{x}_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} \mathbf{x}_j \right\|^2
\]

Choose edge weights to minimize this.
Motivation

Try to get right on coordinate vectors,

$$\mathbf{x}_i = (x^1_i, \ldots, x^d_i)$$

Sum of squares of errors over d coordinate vectors

$$\sum_{k=1}^{d} \sum_{i=1}^{n} \left( x^k_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x^k_j \right)^2$$

$$= \sum_{i=1}^{n} \left\| \mathbf{x}_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} \mathbf{x}_j \right\|^2$$

Choose edge weights to minimize this.

But, leads to a non-convex problem, and...
Motivation, tricky example

Consider minimizing

$$\sum_{i=1}^{n} \left\| x_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x_j \right\|^2$$
Motivation, tricky example

Consider minimizing

\[ \sum_{i=1}^{n} \left\| x_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x_j \right\|^2 \]
Motivation, tricky example

Consider minimizing

\[ \sum_{i=1}^{n} \left\| x_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x_j \right\|^2 \]
Motivation, tricky example

Consider minimizing

\[ \sum_{i=1}^{n} \left\| x_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x_j \right\|^2 \]
Motivation, tricky example

Consider minimizing 

$$\sum_{i=1}^{n} \left\| x_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x_j \right\|^2$$

Just better as $\epsilon$ goes to 0
Motivation, fixing bad example

Sum squares of errors, by weighted degree

\[ \sum_{i=1}^{n} \left\| \mathbf{x}_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} \mathbf{x}_j \right\|^2 \]

\[ \sum_{i=1}^{n} \left\| d_i \mathbf{x}_i - \sum_{j \sim i} w_{i,j} \mathbf{x}_j \right\|^2 \]
Motivation, fixing bad example

Sum squares of errors, by weighted degree

\[
\sum_{i=1}^{n} \left\| x_i - \frac{1}{d_i} \sum_{j \sim i} w_{i,j} x_j \right\|^2 \quad \rightarrow \quad \sum_{i=1}^{n} \left\| d_i x_i - \sum_{j \sim i} w_{i,j} x_j \right\|^2
\]

To avoid degeneracy, require either

\[ d_i = \sum_{j \sim i} w_{i,j} \geq 1 \quad \text{(hard graph)} \]

or

\[ \sum_{i:d_i<1} (1 - d_i)^2 \leq \alpha n \quad \text{($\alpha$-soft graph)} \]
Example

Random points in 2 dimensions (0.1-soft)
Example

Two Clusters (0.1-soft)
Example
(0.1-soft)
Unique?

Not necessarily:
Not unique on highly symmetric point sets.
(will see why later)

But, almost always unique.

Conjecture:
Unique with probability 1 after infinitesimally small random perturbation.
Related to

Locally linear embeddings [Roweis-Saul ‘
Chooses allowable neighbors, then
chooses asymmetric weights to minimize
the same objective function.

K-means
If restrict graph to be a union of complete
graphs, one for each cluster, are minimizing
same same objective function.
In terms of the graph Laplacian

\[ L = D - A \]

where \( A \) is the weighted adjacency matrix, and \( D \) is the diagonal matrix of degrees.
In terms of the graph Laplacian

\[ L = D - A \]

where \( A \) is the weighted adjacency matrix, and \( D \) is the diagonal matrix of degrees.

\[
\sum_{i=1}^{n} \left\| d_i x_i - \sum_{j \sim i} w_{i,j} x_j \right\|^2 = \| LX \|^2_F
\]

\( L \) is linear in edge weights, so \( \| LX \|^2_F \) is quadratic

(term rejected was \( \| D^{-1} LX \|^2_F \))
Sparse Solution

Can fix \( LX \), or difference vectors

\[
\Delta_i = d_i x_i - \sum_{j \sim i} w_{i,j} x_j = \sum_{j \sim i} w_{i,j} (x_i - x_j)
\]

and weighted degrees: \( d = |U| w \)

Imposing only \( n(d+1) \) constraints on \( w \).

If more than \( n(d+1) \) non-zero entries, can make one zero, holding all these fixed, keeping \( w \) non-negative.
Sparsity

Theorem:
For $n$ vectors in $d$ dimensions,
always is a graph with at most $(d+1)n$ edges,
average degree $2(d+1)$.

Theorem (later)
For $n$ vectors in any dimension,
is an $\varepsilon$-approx solution with
average degree $O(1/\varepsilon^2)$. 
## Degrees on real data

<table>
<thead>
<tr>
<th>Data set</th>
<th>Type</th>
<th>n</th>
<th>dim</th>
<th>ave deg (hard)</th>
<th>ave deg (soft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>abalone</td>
<td>regression</td>
<td>4177</td>
<td>8</td>
<td>13.1</td>
<td>12.7</td>
</tr>
<tr>
<td>glass</td>
<td>6 classes</td>
<td>214</td>
<td>9</td>
<td>8.9</td>
<td>8.6</td>
</tr>
<tr>
<td>heart</td>
<td>2 classes</td>
<td>270</td>
<td>13</td>
<td>11.1</td>
<td>11</td>
</tr>
<tr>
<td>housing</td>
<td>regression</td>
<td>506</td>
<td>13</td>
<td>8.8</td>
<td>10.1</td>
</tr>
<tr>
<td>ionosphere</td>
<td>2 classes</td>
<td>351</td>
<td>34</td>
<td>13</td>
<td>11.9</td>
</tr>
<tr>
<td>iris</td>
<td>3 classes</td>
<td>150</td>
<td>4</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>machine</td>
<td>regression</td>
<td>209</td>
<td>6</td>
<td>8.9</td>
<td>8</td>
</tr>
<tr>
<td>mpg</td>
<td>regression</td>
<td>392</td>
<td>7</td>
<td>7.6</td>
<td>8.8</td>
</tr>
<tr>
<td>pima</td>
<td>2 classes</td>
<td>768</td>
<td>8</td>
<td>11.2</td>
<td>10.9</td>
</tr>
<tr>
<td>sonar</td>
<td>2 classes</td>
<td>208</td>
<td>60</td>
<td>12.7</td>
<td>13</td>
</tr>
<tr>
<td>vehicle</td>
<td>4 classes</td>
<td>846</td>
<td>18</td>
<td>11.8</td>
<td>11.5</td>
</tr>
<tr>
<td>vowel</td>
<td>11 classes</td>
<td>990</td>
<td>10</td>
<td>10.4</td>
<td>10.1</td>
</tr>
<tr>
<td>wine</td>
<td>3 classes</td>
<td>178</td>
<td>13</td>
<td>9.1</td>
<td>9.9</td>
</tr>
</tbody>
</table>
# Degrees on real data

<table>
<thead>
<tr>
<th>Data set</th>
<th>Type</th>
<th>n</th>
<th>dim</th>
<th>ave deg (hard)</th>
<th>ave deg (soft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>abalone</td>
<td>regression</td>
<td>4177</td>
<td>8</td>
<td>13.1</td>
<td>12.7</td>
</tr>
<tr>
<td>glass</td>
<td>6 classes</td>
<td>214</td>
<td>9</td>
<td>8.9</td>
<td>8.6</td>
</tr>
<tr>
<td>heart</td>
<td>2 classes</td>
<td>270</td>
<td>13</td>
<td>11.1</td>
<td>11</td>
</tr>
<tr>
<td>housing</td>
<td>regression</td>
<td>506</td>
<td>13</td>
<td>8.8</td>
<td>10.1</td>
</tr>
<tr>
<td>ionosphere</td>
<td>2 classes</td>
<td>351</td>
<td></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>iris</td>
<td>3 classes</td>
<td>150</td>
<td>4</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>machine</td>
<td>regression</td>
<td>209</td>
<td>6</td>
<td>8.9</td>
<td>8</td>
</tr>
<tr>
<td>mpg</td>
<td>regression</td>
<td>392</td>
<td>7</td>
<td>7.6</td>
<td>8.8</td>
</tr>
<tr>
<td>pima</td>
<td>2 classes</td>
<td>768</td>
<td>8</td>
<td>11.2</td>
<td>10.9</td>
</tr>
<tr>
<td>sonar</td>
<td>2 classes</td>
<td>208</td>
<td>60</td>
<td>12.7</td>
<td>13</td>
</tr>
<tr>
<td>vehicle</td>
<td>4 classes</td>
<td>846</td>
<td>18</td>
<td>11.8</td>
<td>11.5</td>
</tr>
<tr>
<td>vowel</td>
<td>11 classes</td>
<td>990</td>
<td>10</td>
<td>10.4</td>
<td>10.1</td>
</tr>
<tr>
<td>wine</td>
<td>3 classes</td>
<td>178</td>
<td>13</td>
<td>9.1</td>
<td>9.9</td>
</tr>
</tbody>
</table>
Example for which graphs are not unique

Consider set of vectors in \( \{0,1\}^d \) of even parity.

Sum of any two solutions is a solution, so sum over all isomorphisms of the point set

Every vertex gets degree at least \( \binom{d}{2} \)

By previous theorem, is a solution of average degree at most \( 2(d + 1) \)

Cannot have symmetry and sparsity
Planarity

For every set of vectors in 2 dimensions, there is a hard and $\alpha$-soft graph that is planar

no crossings
Planarity

For every set of vectors in 2 dimensions, there is a hard and $\alpha$-soft graph that is planar.

And, no point contained in a triangle. In general dimension, is a simplicial complex.
Planarity, proof
A graph with maximum sum of weighted degrees has no crossings.

- increase weights outside
- decrease weights on crossing edges

Will fix $LX$ and thus $\|LX\|_F^2$
but increases all weighted degrees
Planarity, proof

\[ z = \alpha_0 x_0 + \alpha_1 x_1, \]
\[ z = \beta_0 y_0 + \beta_1 y_1, \]
\[ \alpha_0 + \alpha_1 = 1 \]
\[ \beta_0 + \beta_1 = 1 \]

Will fix \( LX \) and thus \( \|LX\|^2_F \)
but increases all weighted degrees
Classification and Regression Experiments

Given a set $S$ of vectors with labels, $y_i \in \mathbb{R}$, compute $z$ minimizing

$$\sum_{i \sim j} w_{i,j} (z_i - z_j)^2 = z^T L z$$

subject to $z_i = y_i$, for $i \in S$  \[ZGL \ '03\]

With 2 clusters, use an indicator vector for one. With $k$ clusters, use $k$ indicator vectors.
Classification Experiments

Used 10-fold cross validation

We had no parameters to train.

For other means of choosing graphs, chose their parameters by gridding over choices, and evaluating by leave-one-out cross-validation.

Repeated 100 times
## Classification experiments

<table>
<thead>
<tr>
<th>Data set</th>
<th>n</th>
<th>dim</th>
<th>classes</th>
<th>0.1-soft</th>
<th>knn</th>
<th>thresh</th>
</tr>
</thead>
<tbody>
<tr>
<td>glass</td>
<td>214</td>
<td>9</td>
<td>6</td>
<td>28.3</td>
<td>26.92</td>
<td>33.3</td>
</tr>
<tr>
<td>heart</td>
<td>270</td>
<td>13</td>
<td>2</td>
<td>17.81</td>
<td>16.05</td>
<td>16.1</td>
</tr>
<tr>
<td>ionosphere</td>
<td>351</td>
<td>34</td>
<td>2</td>
<td>5.57</td>
<td>18.5</td>
<td>6.34</td>
</tr>
<tr>
<td>iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
<td>4.21</td>
<td>4.46</td>
<td>6.2</td>
</tr>
<tr>
<td>pima</td>
<td>768</td>
<td>8</td>
<td>2</td>
<td>26.61</td>
<td>24.54</td>
<td>26.45</td>
</tr>
<tr>
<td>sonar</td>
<td>208</td>
<td>60</td>
<td>2</td>
<td>8.64</td>
<td>13.8</td>
<td>14.94</td>
</tr>
<tr>
<td>vehicle</td>
<td>846</td>
<td>18</td>
<td>4</td>
<td>22.47</td>
<td>27.7</td>
<td>29.98</td>
</tr>
<tr>
<td>vowel</td>
<td>990</td>
<td>10</td>
<td>11</td>
<td>0.95</td>
<td>2.62</td>
<td>0.98</td>
</tr>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>3</td>
<td>2.62</td>
<td>2.86</td>
<td>3.64</td>
</tr>
</tbody>
</table>

tried unweighted, and exponential weights varying $\sigma$
varied $k$ for knn, and $\delta$ for threshold
## Classification experiments

<table>
<thead>
<tr>
<th>Data set</th>
<th>n</th>
<th>dim</th>
<th>classes</th>
<th>0.1-soft</th>
<th>knn</th>
<th>thresh</th>
<th>libsvm</th>
</tr>
</thead>
<tbody>
<tr>
<td>glass</td>
<td>214</td>
<td>9</td>
<td>6</td>
<td>28.3</td>
<td>26.92</td>
<td>33.3</td>
<td>31.44</td>
</tr>
<tr>
<td>heart</td>
<td>270</td>
<td>13</td>
<td>2</td>
<td>17.81</td>
<td>16.05</td>
<td>16.1</td>
<td>17.01</td>
</tr>
<tr>
<td>ionosphere</td>
<td>351</td>
<td>34</td>
<td>2</td>
<td>5.57</td>
<td>18.5</td>
<td>6.34</td>
<td>6.2</td>
</tr>
<tr>
<td>iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
<td>4.21</td>
<td>4.46</td>
<td>6.2</td>
<td>3.87</td>
</tr>
<tr>
<td>pima</td>
<td>768</td>
<td>8</td>
<td>2</td>
<td>26.61</td>
<td>24.54</td>
<td>26.45</td>
<td>23.24</td>
</tr>
<tr>
<td>sonar</td>
<td>208</td>
<td>60</td>
<td>2</td>
<td>8.64</td>
<td>13.8</td>
<td>14.94</td>
<td>11.71</td>
</tr>
<tr>
<td>vehicle</td>
<td>846</td>
<td>18</td>
<td>4</td>
<td>22.47</td>
<td>27.7</td>
<td>29.98</td>
<td>14.87</td>
</tr>
<tr>
<td>vowel</td>
<td>990</td>
<td>10</td>
<td>11</td>
<td>0.95</td>
<td>2.62</td>
<td>0.98</td>
<td>0.64</td>
</tr>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>3</td>
<td>2.62</td>
<td>2.86</td>
<td>3.64</td>
<td>2.57</td>
</tr>
</tbody>
</table>

For libsvm [ChangLin], chose params by 10-fold CV on training dat
## Regression Error

<table>
<thead>
<tr>
<th>Data set</th>
<th>n</th>
<th>dim</th>
<th>hard</th>
<th>0.1-soft</th>
<th>knn</th>
<th>thresh</th>
<th>epsilon-svr</th>
</tr>
</thead>
<tbody>
<tr>
<td>abalone</td>
<td>4177</td>
<td>8</td>
<td>0.479</td>
<td>0.482</td>
<td>0.492</td>
<td>0.657</td>
<td>Too long</td>
</tr>
<tr>
<td>housing</td>
<td>506</td>
<td>13</td>
<td>0.136</td>
<td>0.138</td>
<td>0.224</td>
<td>0.507</td>
<td>0.138</td>
</tr>
<tr>
<td>machine</td>
<td>209</td>
<td>6</td>
<td>0.17</td>
<td>0.185</td>
<td>0.164</td>
<td>0.608</td>
<td>0.394</td>
</tr>
<tr>
<td>mpg</td>
<td>392</td>
<td>7</td>
<td>0.12</td>
<td>0.118</td>
<td>0.137</td>
<td>0.145</td>
<td>0.128</td>
</tr>
</tbody>
</table>

Data normalized to have variance 1. 2-fold CV on abalone, 10-fold otherwise.
## Clustering Experiments (0.1-soft)

<table>
<thead>
<tr>
<th>Data set</th>
<th>n</th>
<th>dim</th>
<th>k</th>
<th>k-means</th>
<th>NJW</th>
<th>Nvec = k</th>
</tr>
</thead>
<tbody>
<tr>
<td>glass</td>
<td>214</td>
<td>9</td>
<td>6</td>
<td>0.41</td>
<td>0.43</td>
<td>0.45</td>
</tr>
<tr>
<td>heart</td>
<td>270</td>
<td>13</td>
<td>2</td>
<td>0.41</td>
<td>0.42</td>
<td>0.19</td>
</tr>
<tr>
<td>ionosphere</td>
<td>351</td>
<td>34</td>
<td>2</td>
<td>0.29</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
<td>0.11</td>
<td>0.33</td>
<td>0.17</td>
</tr>
<tr>
<td>pima</td>
<td>768</td>
<td>8</td>
<td>2</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>sonar</td>
<td>208</td>
<td>60</td>
<td>2</td>
<td>0.45</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td>vehicle</td>
<td>846</td>
<td>18</td>
<td>4</td>
<td>0.55</td>
<td>0.62</td>
<td>0.56</td>
</tr>
<tr>
<td>Vowel</td>
<td>990</td>
<td>10</td>
<td>11</td>
<td>0.66</td>
<td>0.76</td>
<td>0.65</td>
</tr>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>3</td>
<td>0.30</td>
<td>0.33</td>
<td>0.03</td>
</tr>
</tbody>
</table>

NJW: Ng-Jordan-Weiss, as implemented in Spider. Use graph, run k-means in Nvec eigenvectors.
## Clustering Experiments (0.1-soft)

<table>
<thead>
<tr>
<th>Data set</th>
<th>n</th>
<th>dim</th>
<th>k</th>
<th>k-means</th>
<th>NJW</th>
<th>Nvec = k</th>
<th>Nvec chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td>glass</td>
<td>214</td>
<td>9</td>
<td>6</td>
<td>0.41</td>
<td>0.43</td>
<td>0.45</td>
<td>0.45 (12)</td>
</tr>
<tr>
<td>heart</td>
<td>270</td>
<td>13</td>
<td>2</td>
<td>0.41</td>
<td>0.42</td>
<td>0.19</td>
<td>0.19 (2)</td>
</tr>
<tr>
<td>ionosphere</td>
<td>351</td>
<td>34</td>
<td>2</td>
<td>0.29</td>
<td>0.33</td>
<td>0.33</td>
<td>0.09 (15)</td>
</tr>
<tr>
<td>iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
<td>0.11</td>
<td>0.33</td>
<td>0.17</td>
<td>0.09 (8)</td>
</tr>
<tr>
<td>pima</td>
<td>768</td>
<td>8</td>
<td>2</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35 (12)</td>
</tr>
<tr>
<td>sonar</td>
<td>208</td>
<td>60</td>
<td>2</td>
<td>0.45</td>
<td>0.47</td>
<td>0.40</td>
<td>0.35 (35)</td>
</tr>
<tr>
<td>vehicle</td>
<td>846</td>
<td>18</td>
<td>4</td>
<td>0.55</td>
<td>0.62</td>
<td>0.56</td>
<td>0.53 (6)</td>
</tr>
<tr>
<td>Vowel</td>
<td>990</td>
<td>10</td>
<td>11</td>
<td>0.66</td>
<td>0.76</td>
<td>0.65</td>
<td>0.60 (30)</td>
</tr>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>3</td>
<td>0.30</td>
<td>0.33</td>
<td>0.03</td>
<td>0.03 (3)</td>
</tr>
</tbody>
</table>

NJW: Ng, Jordan, Weiss, as implemented in Spider. Use graph, run k-means in Nvec eigenvectors.
Choosing number vectors for spectral

Project $X$ into eigenvectors.
Take vectors until all small coefficient.
(Still fuzzy)
Quadratic Program for Edge Weights

\[ L = U W U^T \]

- Signed vertex-edge adjacency matrix
- Diagonal matrix of edge weights

\[ V \begin{pmatrix} 1 & \mathbf{U} \\ -1 & \end{pmatrix} \]

\[ E \]
Objective function is quadratic

\[ \| L X \|_F^2 = \sum_{k=1}^{d} \| L x^k \|_2^2 = \sum_{k=1}^{d} \| U W U^T x^k \|_2^2 \]

\[ = \sum_{k=1}^{d} \| U W y^k \|_2^2 \]

\[ = \sum_{k=1}^{d} \| U Y^{(k)} w \|_2^2 \]

\[ = \left\| \begin{pmatrix} (U Y^{(1)}) & \vdots & (U Y^{(d)}) \end{pmatrix} w \right\|_2^2 \]

\( w \) is vector of edge weights
Soft graph program

Minimize $\| M w \|^2$

Subject to \[ \sum_{i=1}^{n} (\max(1 - d_i, 0)^2) \leq \alpha n \]

$w \geq 0$

$M = \begin{pmatrix} UY^{(1)} \\ \vdots \\ UY^{(d)} \end{pmatrix}$

$d = |U| w$
Soft graph program

Minimize $\|Mw\|^2$

Subject to $\sum_{i=1}^{n}(\max(1 - d_i, 0)^2) \leq \alpha n \quad d = |U|w$

$w \geq 0$

Minimize $\|Mw\|^2 + \mu \sum_{i=1}^{n}(\max(1 - d_i, 0)^2)$

Searching over $\mu$ until $\sum_{i=1}^{n}(\max(1 - d_i, 0)^2) = \alpha n$
Soft graph program, as NNLSQ

Minimize \[ \| Mw \| ^2 + \mu \sum_{i=1}^{n} (\max(1 - d_i, 0)^2) \]

Minimize \[ \| Mw \| ^2 + \mu \| 1 + s - |U|w \| ^2 \]

\[ w, s \geq 0 \]

A non-negative least-squares problem.
Computing Soft graph quickly

Minimize \[ \| Mw \|^2 + \mu \| 1 + s - |U| w \|^2 \]

\[ w, s \geq 0 \]

Look for solution among a subset of edges

Given solution restricted to subset of edges
check gradient of obj func wrt unused edges
if all negative, finished
otherwise, include some and solve again.
## Computation Time (secs)

<table>
<thead>
<tr>
<th>Data set</th>
<th>Type</th>
<th>n</th>
<th>dim</th>
<th>soft time</th>
<th>hard time</th>
</tr>
</thead>
<tbody>
<tr>
<td>abalone</td>
<td>regression</td>
<td>4177</td>
<td>8</td>
<td>1582</td>
<td>49986</td>
</tr>
<tr>
<td>glass</td>
<td>6 classes</td>
<td>214</td>
<td>9</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>heart</td>
<td>2 classes</td>
<td>270</td>
<td>13</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td>housing</td>
<td>regression</td>
<td>506</td>
<td>13</td>
<td>23</td>
<td>114</td>
</tr>
<tr>
<td>ionosphere</td>
<td>2 classes</td>
<td>351</td>
<td>34</td>
<td>72</td>
<td>148</td>
</tr>
<tr>
<td>iris</td>
<td>3 classes</td>
<td>150</td>
<td>4</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>machine</td>
<td>regression</td>
<td>209</td>
<td>6</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>mpg</td>
<td>regression</td>
<td>392</td>
<td>7</td>
<td>16</td>
<td>43</td>
</tr>
<tr>
<td>pima</td>
<td>2 classes</td>
<td>768</td>
<td>8</td>
<td>111</td>
<td>529</td>
</tr>
<tr>
<td>sonar</td>
<td>2 classes</td>
<td>208</td>
<td>60</td>
<td>38</td>
<td>50</td>
</tr>
<tr>
<td>vehicle</td>
<td>4 classes</td>
<td>846</td>
<td>18</td>
<td>159</td>
<td>889</td>
</tr>
<tr>
<td>vowel</td>
<td>11 classes</td>
<td>990</td>
<td>10</td>
<td>85</td>
<td>706</td>
</tr>
<tr>
<td>wine</td>
<td>3 classes</td>
<td>178</td>
<td>13</td>
<td>6</td>
<td>15</td>
</tr>
</tbody>
</table>
Approximate Sparse Solutions

For every graph $G$ exists a sparse graph $\tilde{G}$ with average degree at most $16/\epsilon^2$

So that for all $x$,

$$\left| x^T L^2 x - x^T \tilde{L}^2 x \right| \leq \epsilon \sqrt{x^T L x} \| L \| \| x \|$$

If not too small, $\| LX \|_F^2 \approx \| \tilde{L} X \|_F^2$

Leave-one-out regression scores similar
Approximate Sparse Solutions

For every graph $G$ exists a sparse graph $\tilde{G}$ with average degree at most $16/\epsilon^2$

So that for all $x$,

$$\left| x^T L^2 x - x^T \tilde{L}^2 x \right| \leq \epsilon (x^T L x) \|L\| \|x\|$$

Want

$$\left| x^T L^2 x - x^T \tilde{L}^2 x \right| \leq \epsilon x^T L^2 x$$
Sparsification

Theorem [Batson-S-Srivastava ‘09]
For every graph $G$ exists a sparse graph $\tilde{G}$ with at most $4n/\epsilon^2$ edges

Such that for all $x$,

$$(1 - \epsilon)x^T \tilde{L}x \leq x^T Lx \leq (1 + \epsilon)x^t \tilde{L}x$$
Sparsification

Theorem [Batson-S-Srivastava ‘09]
For every graph $G$ exists a sparse graph $\tilde{G}$
with at most $4n/\varepsilon^2$ edges

Such that for all $x$,

$$(1 - \varepsilon)x^T \tilde{L}x \leq x^T Lx \leq (1 + \varepsilon)x^t \tilde{L}x$$

<table>
<thead>
<tr>
<th>Edges</th>
<th>Time</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O(n \log n/\varepsilon^2)$</td>
<td>(linsolve) $\log n/\varepsilon^2$</td>
<td>[S-Srivastava ‘08]</td>
</tr>
<tr>
<td>$O(n \log^{29} n/\varepsilon^2)$</td>
<td>$O(m \log^{17} n/\varepsilon^2)$</td>
<td>[S-Teng ‘04]</td>
</tr>
</tbody>
</table>
Lovasz’s Graphs


Form a matrix $M = P - A$,

- $P$ is positive diagonal
- $A$ is non-negative (adjacency)

$MX = 0$ (in fact, $\text{null}(M) = X$)

$M$ has one negative eigenvalue

Works if points are on convex hull of polytope.

$A$ non-zero only for edges of polytope.
Open Questions

Are there other natural and useful graphs?

Are there interesting ways to modify/parameterize our construction?

Are our graphs connected?
Are our graphs unique (under perturbation)?

Do there exist sparse approximate solutions in all dimensions?

Can one sparsify for the square of the Laplacian?

How to interpolate off the graph, or add points?