Semi-supervised Regression with Temporal Image Sequences

Ling Xie, Miguel Á. Carreira-Perpiñán, and Shawn Newsam

Electrical Engineering & Computer Science

University of California at Merced

ICIP 2010

September 28th, 2010

The Problem

- Semi-supervised regression in which:
  - Have labelled and unlabelled training data.
  - Have knowledge about relation between data.
    - E.g., temporal information.
  - Challenge: how to use unlabelled data to improve prediction model?

Motivating Problem

- Estimating atmospheric visibility using cameras instead of specialized equipment.

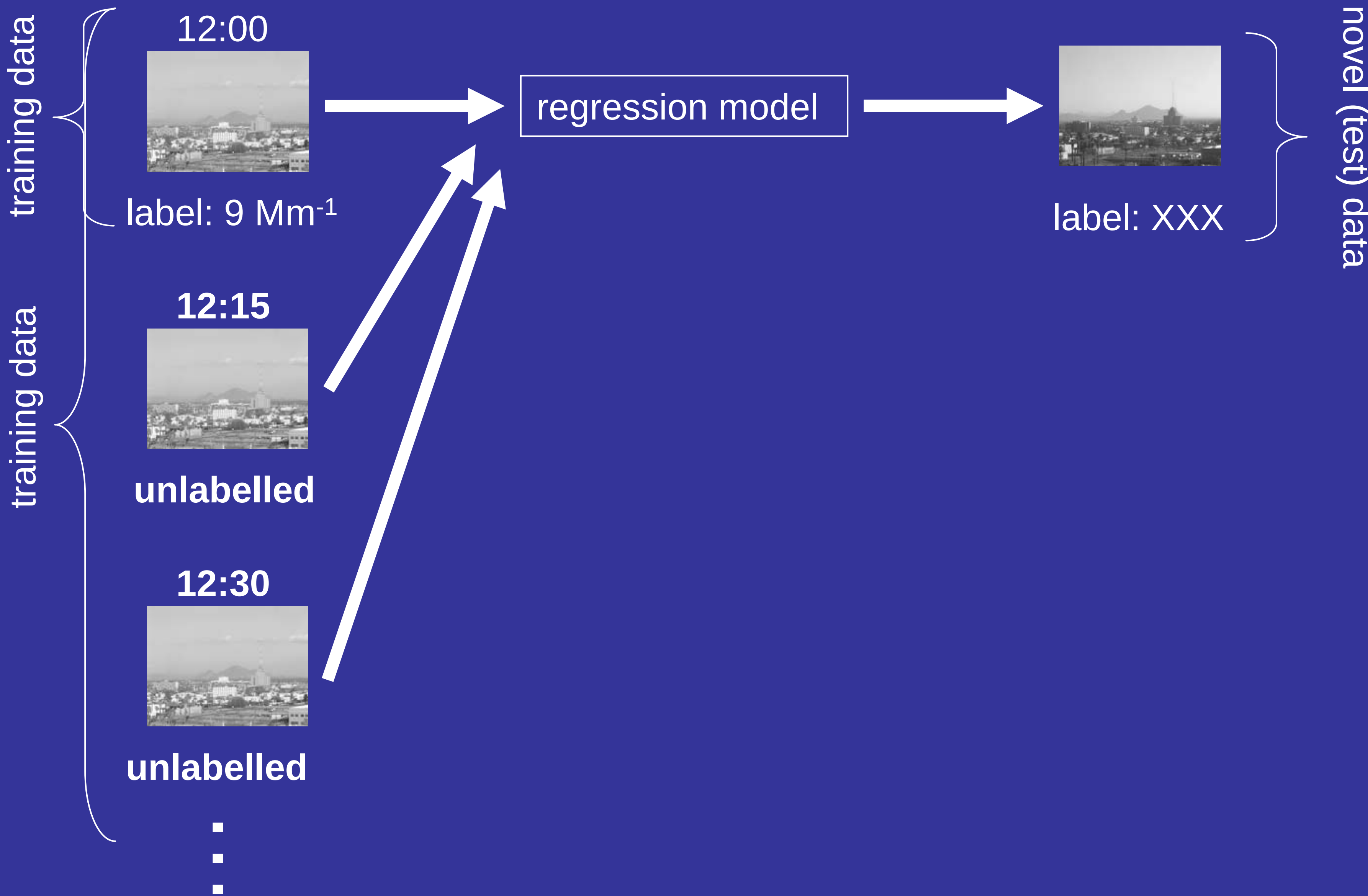
(a) $b_{ext} = 9\ \mathrm{Mm}^{-1}$ (b) $b_{ext} = 28\ \mathrm{Mm}^{-1}$ (c) $b_{ext} = 163\ \mathrm{Mm}^{-1}$

Figure 14: Images demonstrating difference levels of visibility according to measurements of the coefficient of extinction, b_{ext}, as measured using a transmissometer. (a) Good visibility. (b) Moderate visibility. (c) Poor visibility.

Motivating Problem – The Twist

- Training dataset:
  - Images acquired every 15 minutes.
  - Ground truth measurements (b_{ext}) acquired **only every hour**.
- Question: can the unlabelled images be used to learn a better prediction model?
- Semi-supervised learning problem with temporal prior.

Motivating Problem



Approach Summary

- Regularized semi-supervised regression.
- Use temporal information to regularize model.
- Assume: labels (e.g., extinction coefficient) vary slowly with time.
- Similar to but different from standard regularized semi-supervised regression which assumes labels vary slowly with inputs (e.g., images).
- Our approach does not assume labels vary slowly with inputs.

Method

- Training set of N labeled points $\{(x_n, y_n)\}$ where $x_n \in \Re^L$ and $y_n \in \Re$.
- M additional unlabelled points $\{(x_m)\}$.
- All M+N points come as a collection of sequences of the form $(x_1, x_2, x_3, \ldots)$ each of which is only partially labelled with y values.
- We want to learn a regression mapping f that predicts the label y for an input x.

UCMERCED

Method

- Consider a least-squares regression setting with a graph Laplacian regularization:

$$E(\mathbf{f}) = \sum_{n=1}^{N} \left\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\right\|^2 + \gamma_A \left\|\mathbf{f}\right\|_K^2 + \gamma_I \left\|\mathbf{f}\right\|_G^2$$

RKHS norm which encourages smoothness regardless of the data distribution

graph Laplacian which encourages smoothness with respect to the distribution of both the labelled and unlabelled training points

Method

- In our case, the graph Laplacian term reduces to:

$$\|\mathbf{f}\|_{G_t}^2 = \mathbf{f}^T \mathbf{L}_t \mathbf{f} = \sum_{\substack{\text{sequen-}\\ \text{ces } s}} \sum_{n=2}^{N_s} w_{n,n-1}^s \left\| \mathbf{f}(\mathbf{x}_n^s) - \mathbf{f}(\mathbf{x}_{n-1}^s) \right\|^2 \quad (2)$$

- Compare this to the usual graph Laplacian term:

$$\|\mathbf{f}\|_G^2 = \mathbf{f}^T \mathbf{L} \mathbf{f} = \sum_{n \sim m} w_{nm} \left\| \mathbf{f}(\mathbf{x}_n) - \mathbf{f}(\mathbf{x}_m) \right\|^2.$$

Method

- Solution of regularized least squares problem is unique and given by a basis function expansion at each of the labelled and unlabelled points:

$$\mathbf{f}(\mathbf{x}) = \sum_{n=1}^{N+M} \alpha_n K(\mathbf{x}_n, \mathbf{x})$$

- We use Gaussian kernels $K(\cdot, \cdot)$ with width σ.

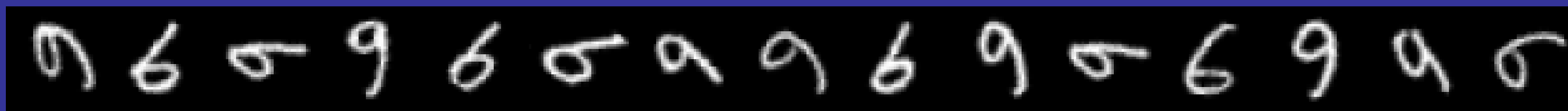
Experiments

- In all experiments, compare:
  - Proposed temporal Laplacian regularized least squares regression (TLapRLSR).
  - Standard Laplacian regularized least squares regression (LapRLSR) where the graph is constructed using k-nearest neighbors in input (image) space.

Experiments

- Image data: rotated MNIST digits.
- Image data: estimating time of day.
- Image data: estimating visibility.

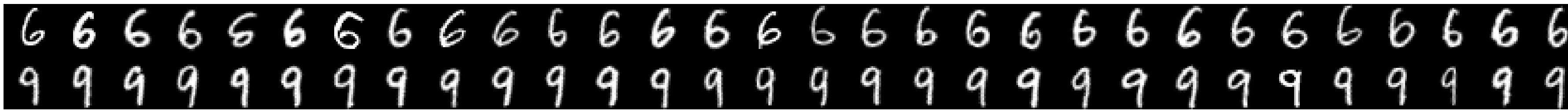
Exp. 1: Rotated MNIST Digits



- What are the angles of these digits?
- Are they sixes at x degrees or nines at x+180 degrees?

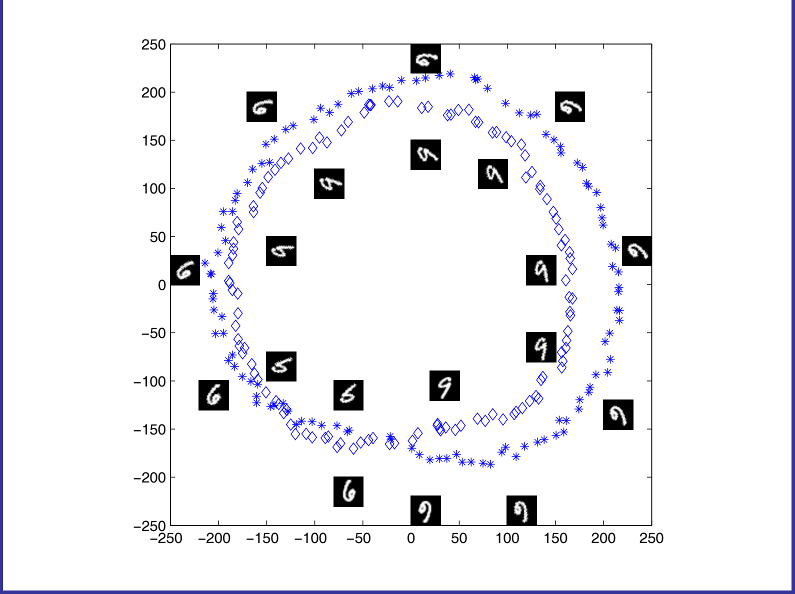
Exp. 1: Rotated MNIST Digits

- Dataset: 30 sixes and 30 nines from MNIST set.



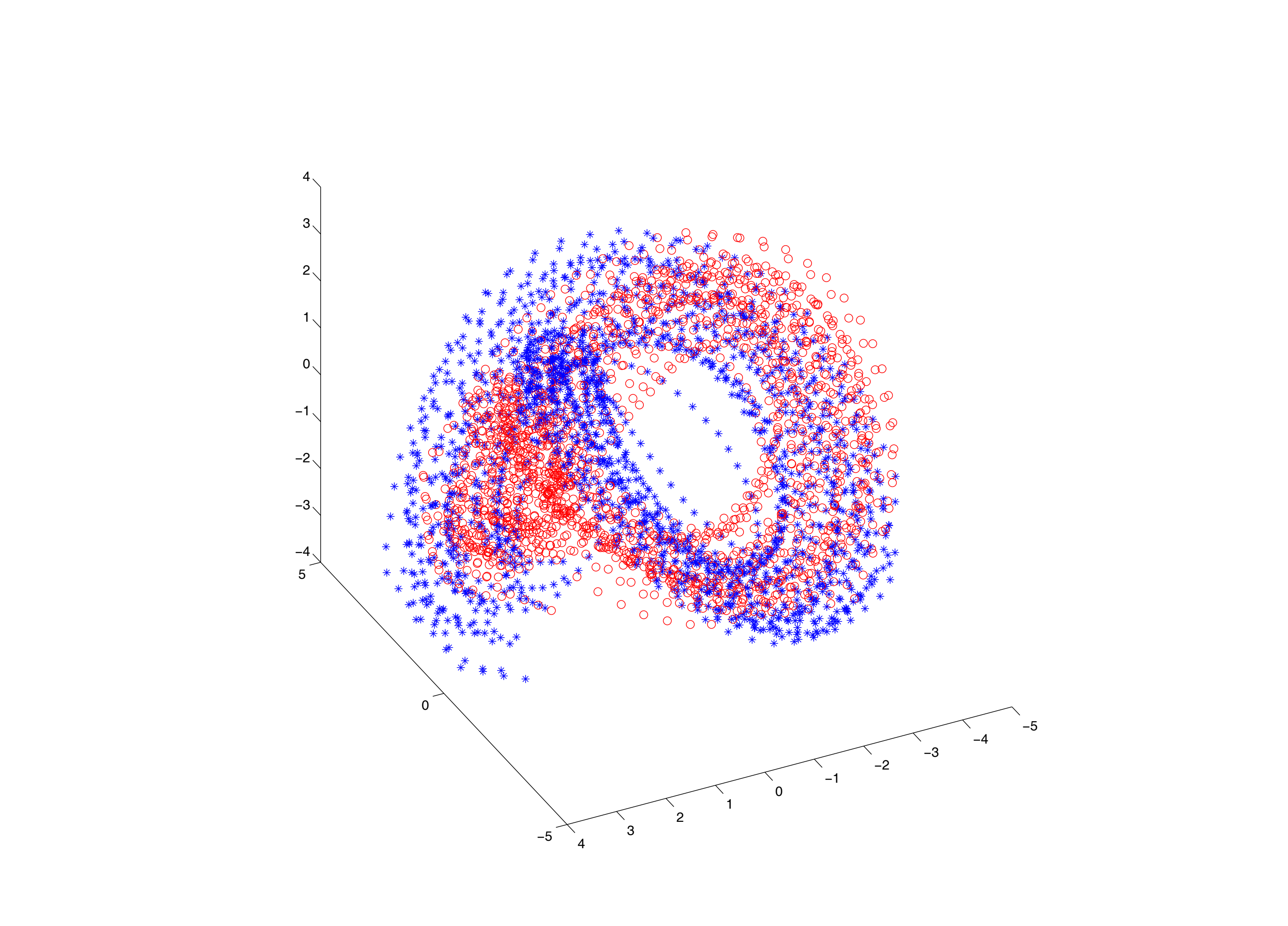
- Each nine is rotated counter-clockwise at one degree intervals from 0 to 180 degrees.
- Each six is rotated counter-clockwise at one degree intervals from 180 to 360 degrees.
- A six might appear very similar to a nine except for a 180 degree phase difference.

Exp. 1: Rotated MNIST Digits



- Images corresponding to a rotated six sequence (asterisks) and a rotated nine sequence (diamonds) projected on the first two PCA components.
- Both sequences start at bottom and go counter-clockwise.
- Sequences are actually (parallel) spirals in the three dimensional PCA space.

E

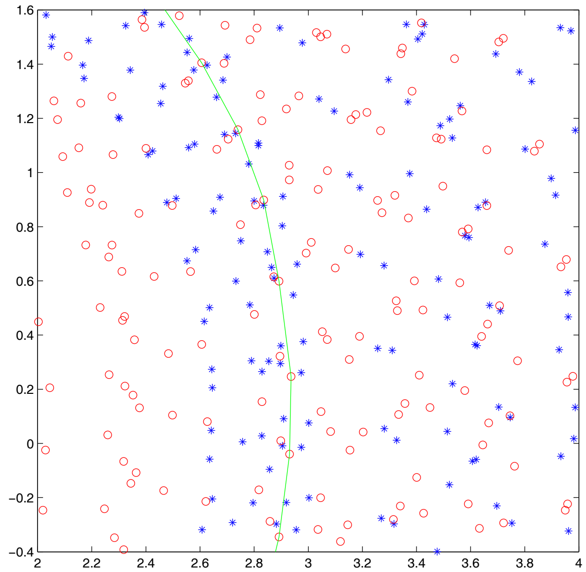


- The rotated sequences of sixes (red) and nines (blue) projected onto the first three PCA components.

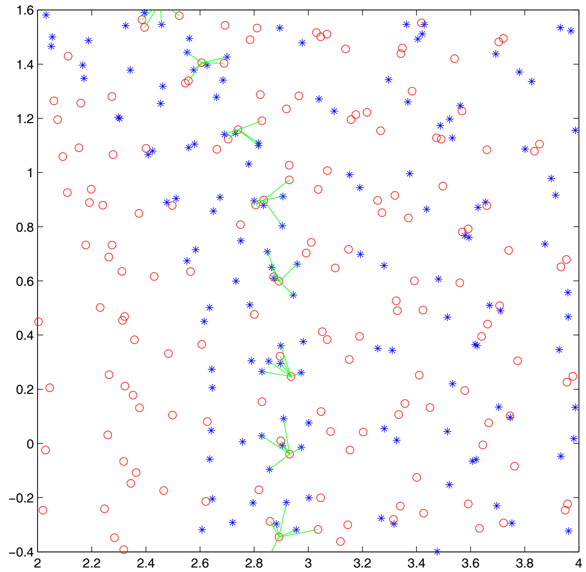
Exp. 1: Rotated MNIST Digits

- The elements of each sequence are assigned in an alternating fashion to training, cross-validation, and evaluation sets.
- A percentage of the training set is labelled (with the rotation angle) at approximately equal spacing with respect to angle.
- The 60 sequences of labelled and unlabelled images in the training set form the “temporal” sequences for TLapRLSR.

Exp. 1: Rotated MNIST Digits



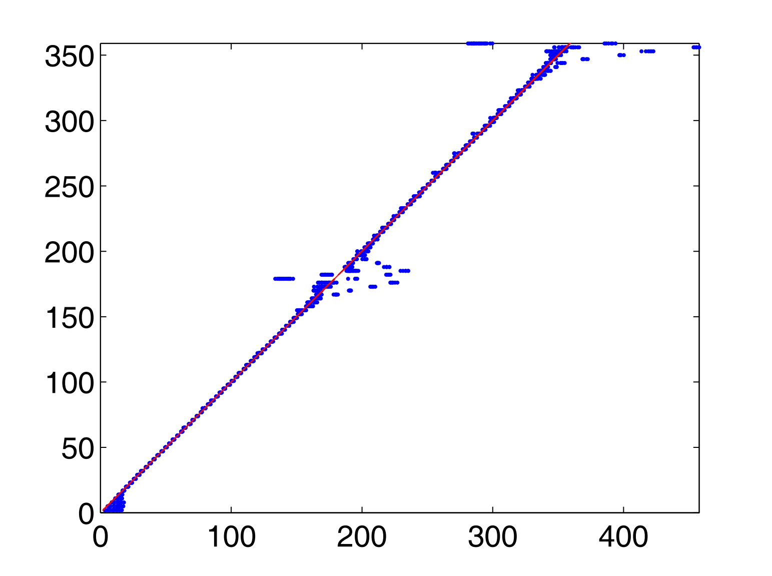
TLapRLSR



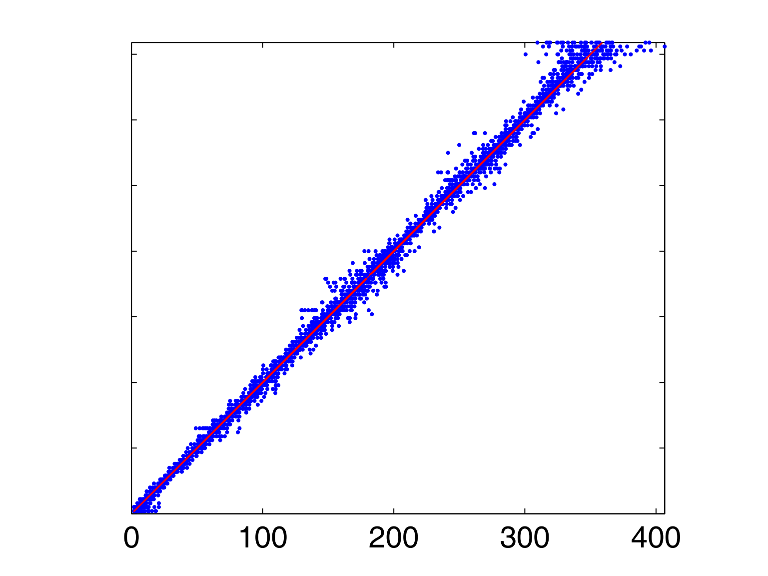
LapRLSR

- Examples of graphs constructed for the rotated digits problem using the sequence information (left) and the standard approach with 6 nearest neighbors (right).
- Circles represent sixes and asterisks represent nines.

Exp. 1: Rotated MNIST Digits



TLapRLSR



LapRLSR

- True (Y-axis) vs. predicted (X-axis) angle on test images for TLapRLSR (left) and LapRLSR (right).
- Nines are plotted from 0 to 180 degrees and sixes from 180 to 360 degrees.
- The diagonal red line represents perfect performance.

Exp. 1: Rotated MNIST Digits

	Feature Space	
	Native	PCA
TLap	92.7 (**2.59**)	**2703 (35.6)**
Lap k=1	189 (8.47)	3330 (40.1)
Lap k=2	32.9 (2.88)	3180 (38.5)
Lap k=3	49.2 (4.26)	3130 (38.0)
Lap k=4	**27.3** (2.87)	3090 (37.5)
Lap k=5	42.7 (3.92)	3080 (37.4)
Lap k=6	41.7 (3.82)	3060 (37.2)
Lap k=7	52.0 (4.42)	3040 (37.1)

- Prediction errors: root-mean square (mean abs).
- Native = image distances computed in raw pixel space.
- PCA = image distances computed in 3 dim. PCA space.

Exp. 2: Estimating Time of Day

- Problem: use an image's visual characteristics to estimate the time of day it was captured.
- Will again construct a training dataset with labelled and unlabelled data.
- The value being predicted, time, varies slowly.
- Two images which appear similar might have been taken at very different times, however. Thus, the feature space might not be smooth with respect to the labels.

Exp. 2: Estimating Time of Day

- Three images taken on the same day.
- The feature space is not necessarily smooth with respect to the label (time).

Exp. 2: Estimating Time of Day

- Dataset:
  - 870 images of outdoor scene acquired from 4:40am to 7:15pm at one minute intervals.
  - Divided into training, cross-validation, and evaluation sets.
  - Approx. 20% of the training images are labelled (with the time of day) at equally spaced time intervals.
- Features:
  - First three PCA components.
  - Raw pixels values after resizing to 64x64.

Exp. 2: Estimating Time of Day

	Feature Space	
	Native	PCA
TLap	**0.111**	**0.568**
Lap k=1	0.346	1.07
Lap k=2	0.330	1.02
Lap k=3	0.329	1.03
Lap k=4	0.332	0.967
Lap k=5	0.345	0.930
Lap k=6	0.348	0.930
Lap k=7	0.349	0.995

- Prediction errors: root-mean square.
- Native = image distances computed in raw pixel space.
- PCA = image distances computed in 3 dim. PCA space.

Exp. 3: Estimating Visibility

- Estimating atmospheric visibility using cameras instead of specialized equipment.

(a) $b_{ext} = 9\ \mathrm{Mm}^{-1}$ (b) $b_{ext} = 28\ \mathrm{Mm}^{-1}$ (c) $b_{ext} = 163\ \mathrm{Mm}^{-1}$

Figure 14: Images demonstrating difference levels of visibility according to measurements of the coefficient of extinction, b_{ext}, as measured using a transmissometer. (a) Good visibility. (b) Moderate visibility. (c) Poor visibility.

Exp. 3: Estimating Visibility

- Dataset:
  - 457 images taken every 15 minutes during daylight hours over a period of two weeks.
  - Have ground truth labels from transmissometer every one hour (123 labelled images).
  - Set aside every other labelled image as test set.
  - Remaining images are labelled and unlabelled training images.
- Features:
  - First three PCA components.
  - Raw pixels values after resizing to 64x64.

Exp. 3: Estimating Visibility

	Feature Space	
	Native	PCA
TLap	**116.7**	**159.2**
Lap k=1	120.8	161.6
Lap k=2	119.0	160.0
Lap k=3	119.5	159.8
Lap k=4	120.4	159.9
Lap k=5	119.6	159.9
Lap k=6	118.3	160.1
Lap k=7	118.0	160.4

- Prediction errors: root-mean square.
- Native = image distances computed in raw pixel space.
- PCA = image distances computed in 3 dim. PCA space.

Conclusion

- Proposed approach, temporal Laplacian regularized least squares regression (TLapRLSR), is shown to outperform standard graph Laplacian regression using k-nearest neighbors.
- Supported by quantitative results for three problems.

Conclusion

- Approach requires:
  - Prior knowledge on how value of interest (angle, time, visibility) varies.
  - Relation of unlabelled to labelled data (such as sequences).
- Currently applied to temporal sequences.
- But other types of structured data is possible such as spatial (geographic) data.

Conclusion

- Please see the paper for:
  - Additional technical details.
  - Sensitivity and convexity of parameter values.
  - The effect of the percent of labelled points.

Acknowledgements:

- Carreira-Perpiñán funded by NSF award IIS-0754089 (CAREER).
- Newsam and Xie funded in part by the Center for Information Technology Research in the Interest of Society (CITRIS) and a UC TSR&TP Atmospheric Aerosols and Health Graduate Fellowship.
- We would like to thank the Arizona Department of Environmental Quality and Air Resource Specialists, Inc., for providing the Phoenix image and transmissometer data.

Thank you! and questions?