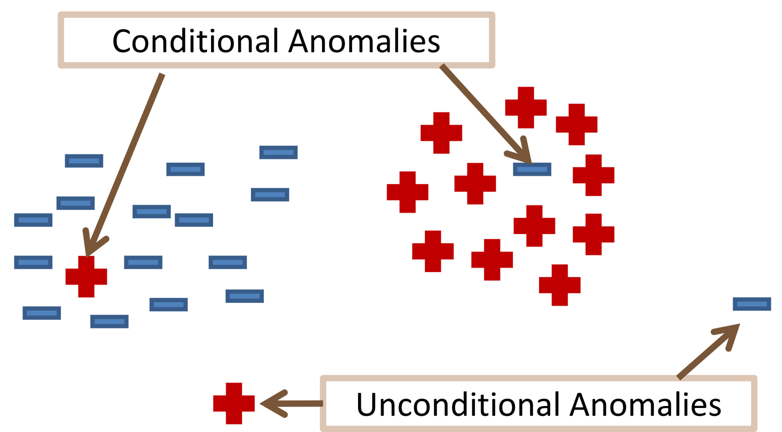
Soft Harmonic Functions For Anomaly Detection

MICHAL VALKO, HAMED VALIZADEGAN, BRANISLAV KVETON, MILOS HAUSKRECHT

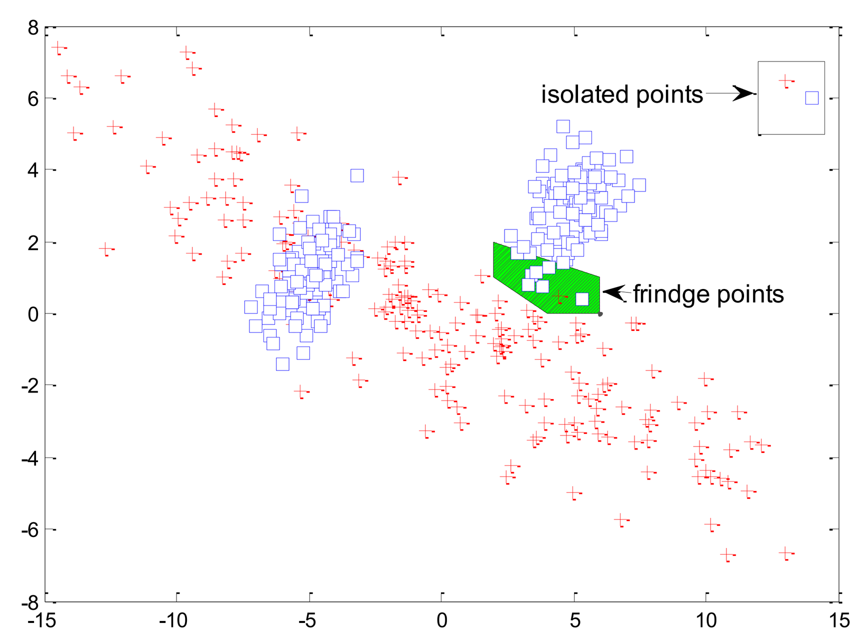
Motivation



- traditionally, anomalies in the data
- we want to detect **anomalies in responses**
- conditioning on the remaining features/covariates
- useful for medical application
  - anomalies in lab orders and medication
  - budget control

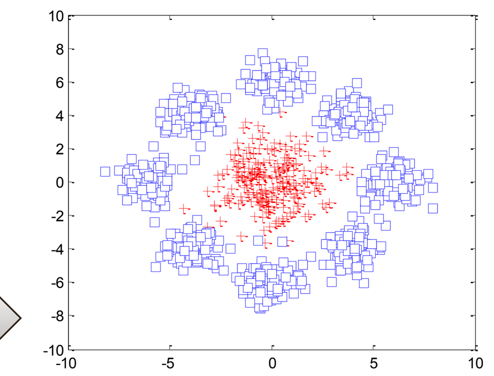
Challenges

- underlying density is often unknown
- high-dimensional and non-linear data
- fringe points (on the support boundary)
- isolated points (unconditional outliers)



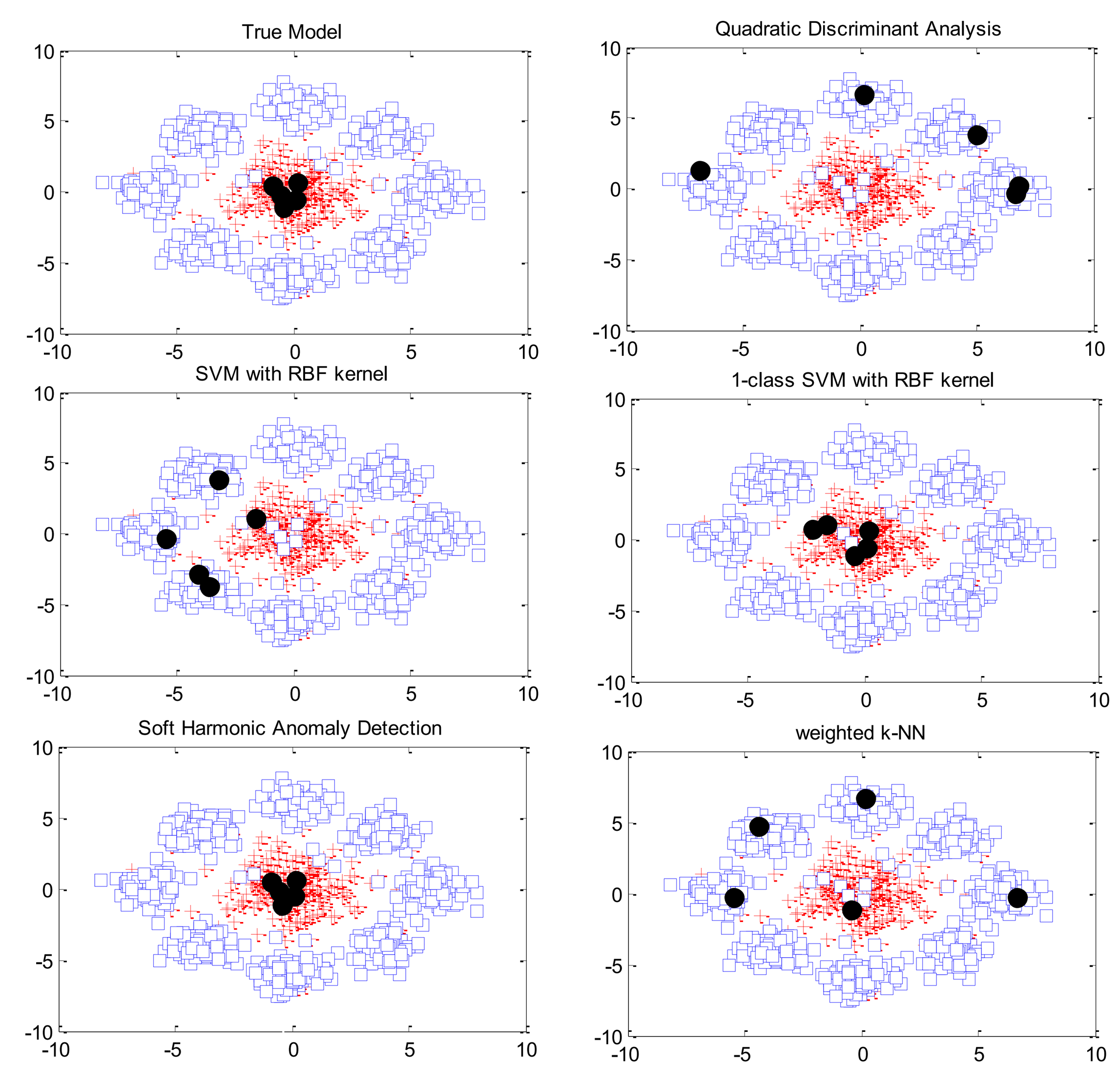
Comparison

- evaluation of conditional anomaly methods is very challenging
- synthetic data with known distribution
- flip 3% of the labels
- compare how the anomaly score agrees with true score



QDA	61.0% (1.2)
SVM	46.1% (3.1)
1-class SVM	64.7% (0.7)
[redacted]	
wk–NN	61.4% (2.1)

Top 5 best scoring anomalies for different methods on the synthetic dataset

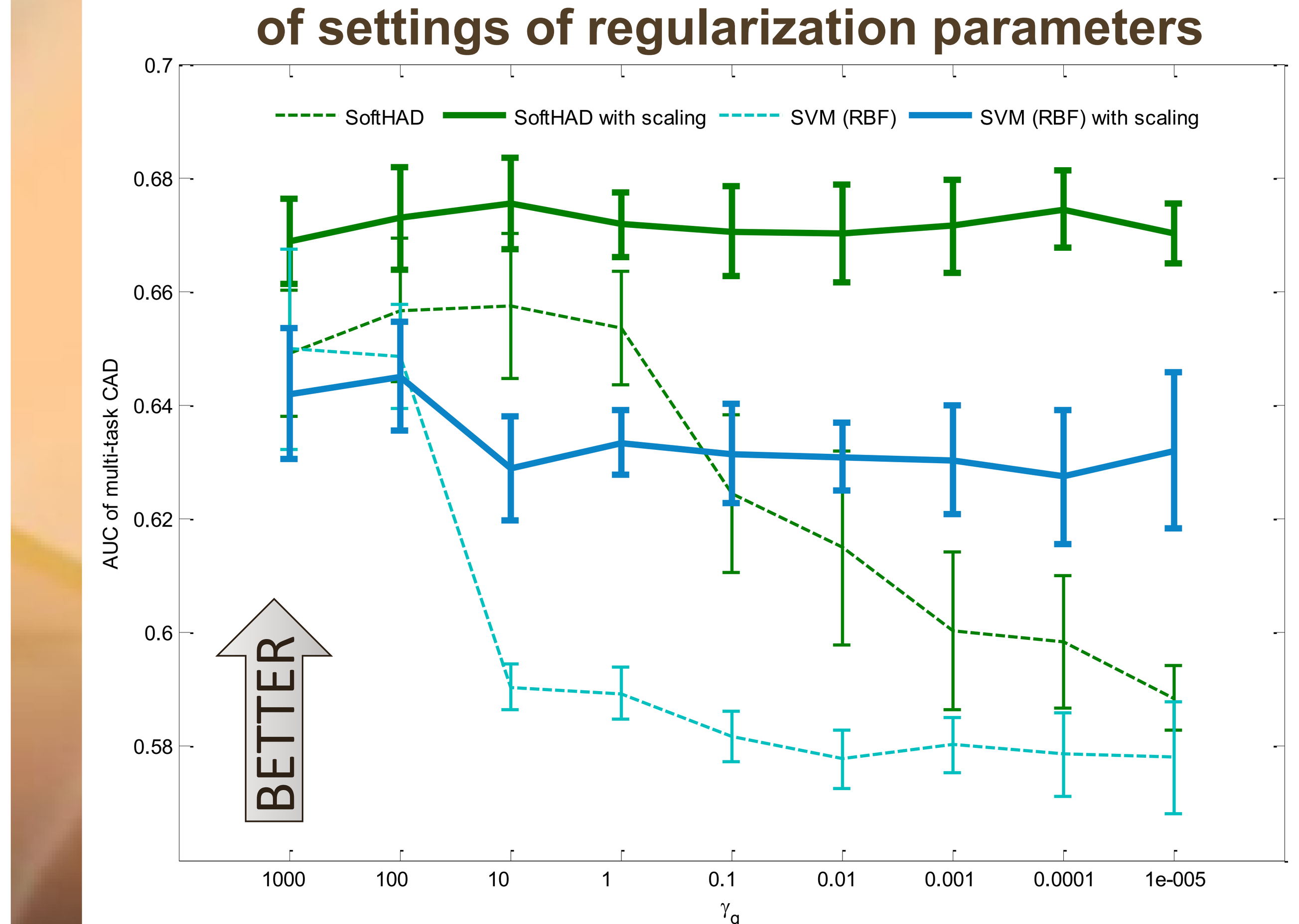


our method

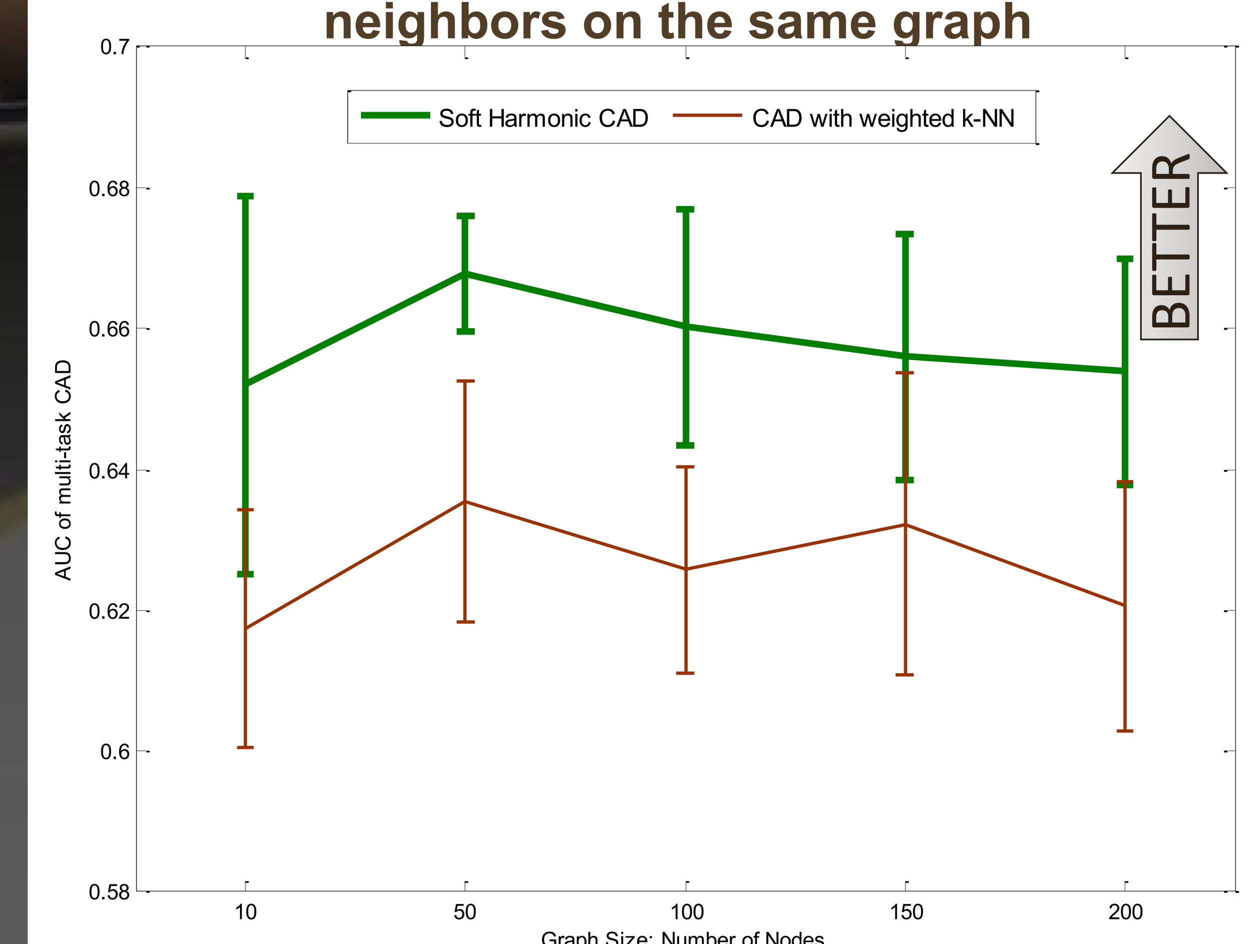
Results

- medical health records (UMPC)
- 4486 patients (50K instances, 9K features)
- 749 laboratory tests or medication orders
- 222 instances evaluated
- panel of expert clinicians (3 per instance)
- evaluation metric: area under ROC

Outperforming SVM method over the range of settings of regularization parameters



Outperforming standard weighted nearest neighbors on the same graph



Background

Goal: Conditional Anomaly Detection

- detect anomalous decisions
- robust to traditional outliers

Problem statement (★): For a dataset $(\mathbf{x}_i, \mathbf{y}_i)_{i=1}^n$ find pairs of $(\mathbf{x}_i, \mathbf{y}_i)$ such that $P(\mathbf{y} \neq \mathbf{y}_i | \mathbf{x}_i)$ is high.

Alternative methods

- class outlier approach
  - take traditional anomaly detection method
  - detect anomalies within the same class
  - cons: ignores the other classes
- discriminative approach
  - sensitive to fringe and isolated points

Our method takes all classes into account and uses **regularization** to avoid unwanted behavior.

Algorithm

- graph-based representation

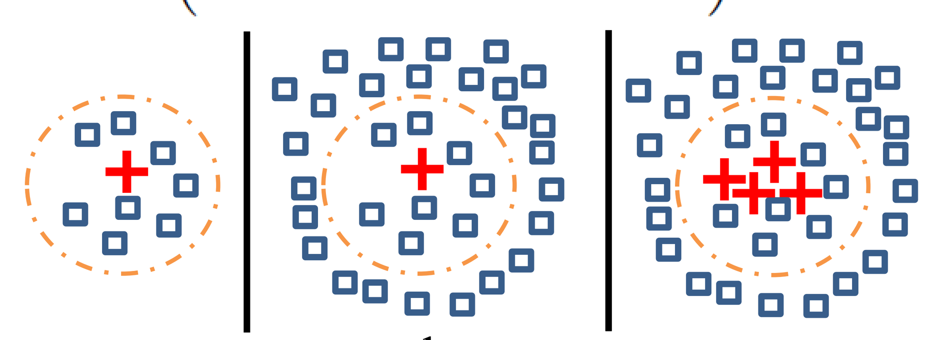
$$w_{ij} = \exp\left[-\left(\|\mathbf{x}_i - \mathbf{x}_j\|_{2,\psi}^2\right) / \sigma^2\right]$$

- label propagation on graph

$$\boldsymbol{\ell}^\star = \min_{\boldsymbol{\ell} \in \mathbb{R}^n} (\boldsymbol{\ell} - \mathbf{y})^\mathsf{T} C (\boldsymbol{\ell} - \mathbf{y}) + \boldsymbol{\ell}^\mathsf{T} K \boldsymbol{\ell}$$

- regularization to prevent unwanted anomalies
- checking for inconsistencies

$$\ell^\star = \left((c_l I)^{-1} (\mathcal{L}(W) + \gamma_g) + I\right)^{-1} \mathbf{y}$$



- addressing computational complexity
  - create a backbone graph
  - make the calculation on a smaller graph

Contributions

- non-parametric and graph-based method for conditional anomaly detection
- takes advantage of the data structure
- important application for medical data
- robust to fringe and isolated points