Identifying Hidden Contexts

Indrė Žliobaitė

TU Eindhoven
2010 08 20

This talk is about

- hidden context / dependencies in the data

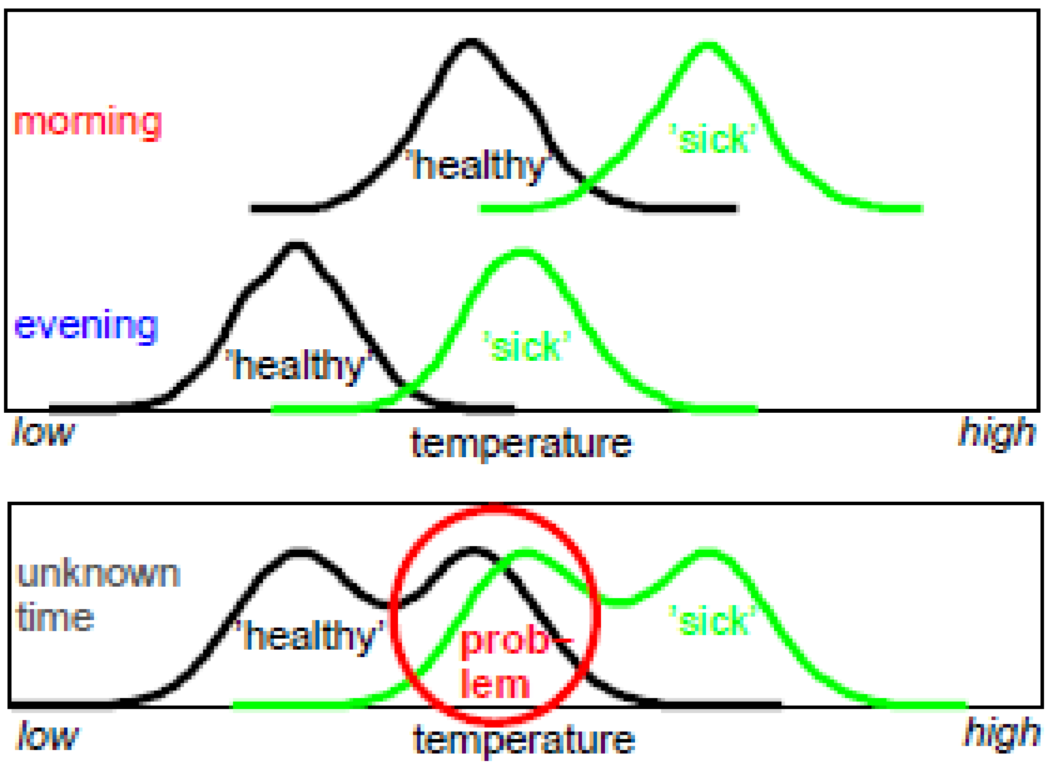
- What is context?
- Why identify hidden context?
- Data transformation (rotation) for identifying context
- Outlook

What is context?

- Supervised learning: (X,y) data and labels
- (X,y,z) data, labels and context

What is context?

- Supervised learning: (X,y) data and labels
- (X,y,z) data, labels and context



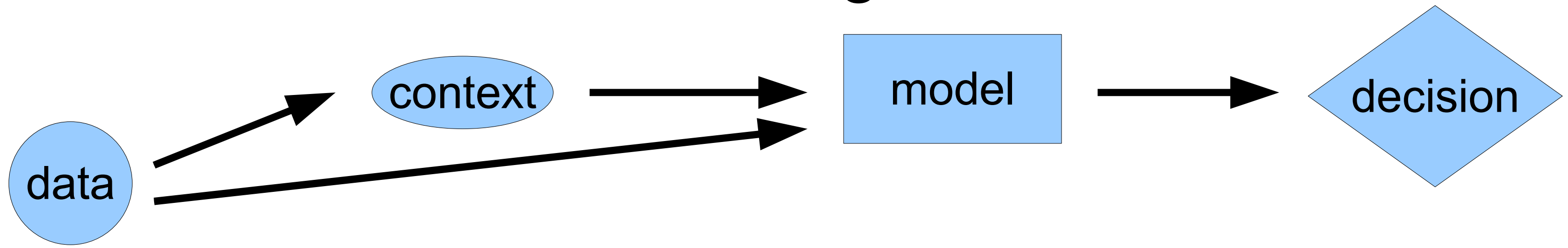
- time does not explain the label $P(y|z) = P(y)$
- but time helps to explain the label using temperature $P(y|X,z) \neq P(y|X)$

Why identifying context?

- To improve classification accuracy
  - two level decision making in static scenarios
  - dynamic (evolving) scenarios
- Better understand the data
- Identifying contexts can be seen as data preprocessing (filter) step

Why identifying context?

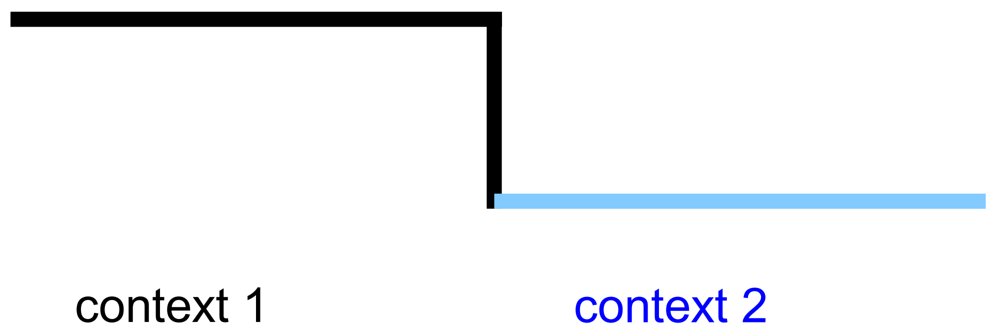
- To improve classification accuracy
  - two level decision making in static scenarios



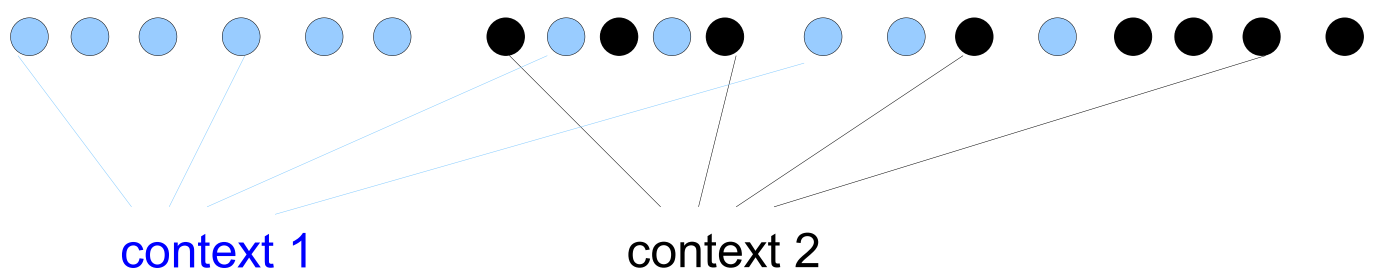
  - dynamic (evolving) scenarios
- Better understand the data
- Identifying contexts can be seen as data preprocessing (filter) step

What about concept drift?

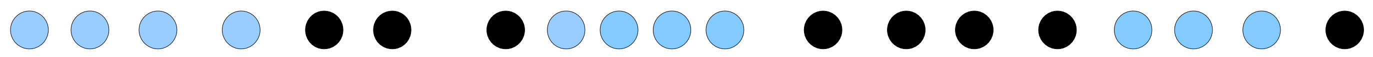
Time
context



Space
context



Reoccuring
context

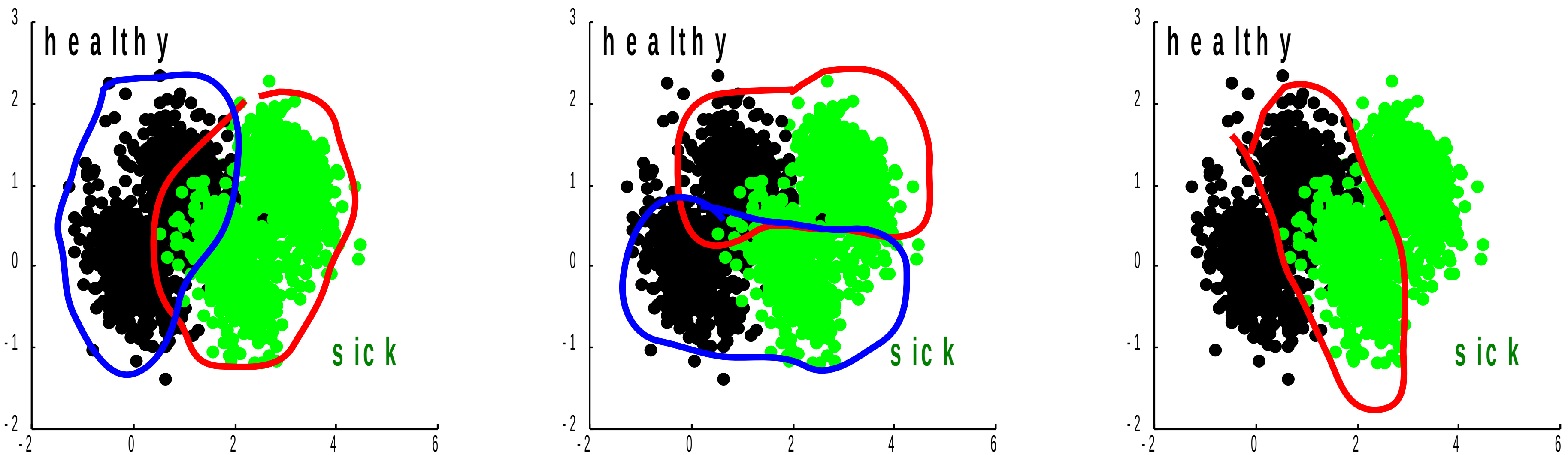


Identifying hidden context

- Grouping (clustering) the data (X)
- But clases (y) instead of contexts (z) might be captured
- Or a mix of classes-contexts

Identifying hidden context

- Grouping (clustering) the data (X)
- But clases (y) instead of contexts (z) might be captured
- Or a mix of classes-contexts



Proposed Approach

- 'Hide' class label information from the data
- i.e. transform the data ($X \rightarrow X'$) so that it is not correlated with the label (y)
- then cluster the transformation X'
- record the contexts (z)
- add context categories to the original dataset (X,y,z)

How to 'hide' class label information

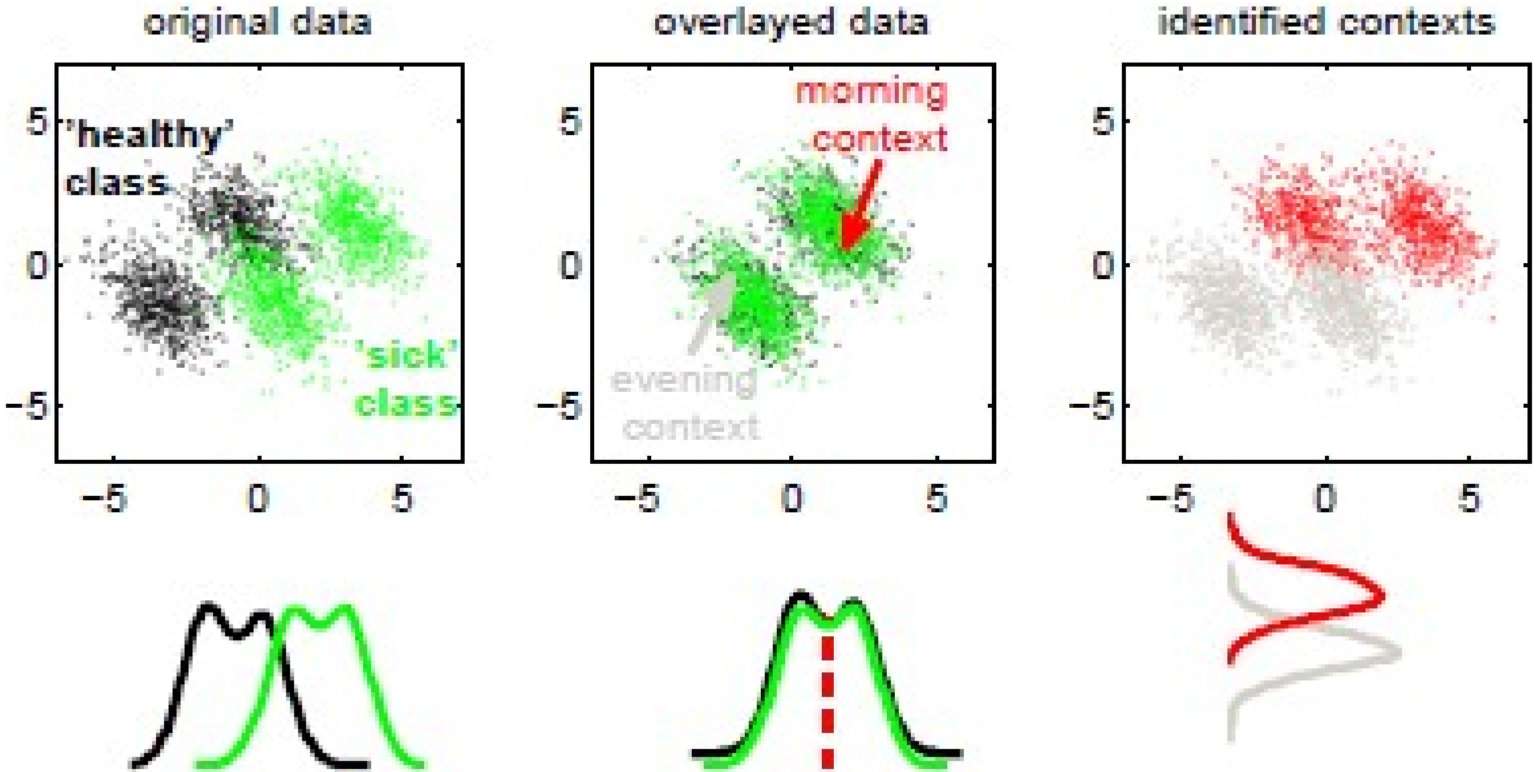
- Three apporaches:
  - Overlay
  - Projections
  - Fature underselection

Overlay

- Normalize each class separately to *zero* mean
- $X' = \{X^{(I)} - mean(X^{(I)}) \cup X^{(II)} - mean(X^{(II)})\}$
- Cluster X'

Overlay

- Normalize each class separately to *zero* mean
- $X' = \{X^{(I)} - \text{mean}(X^{(I)}) \cup X^{(II)} - \text{mean}(X^{(II)})\}$
- Cluster X'



Projection

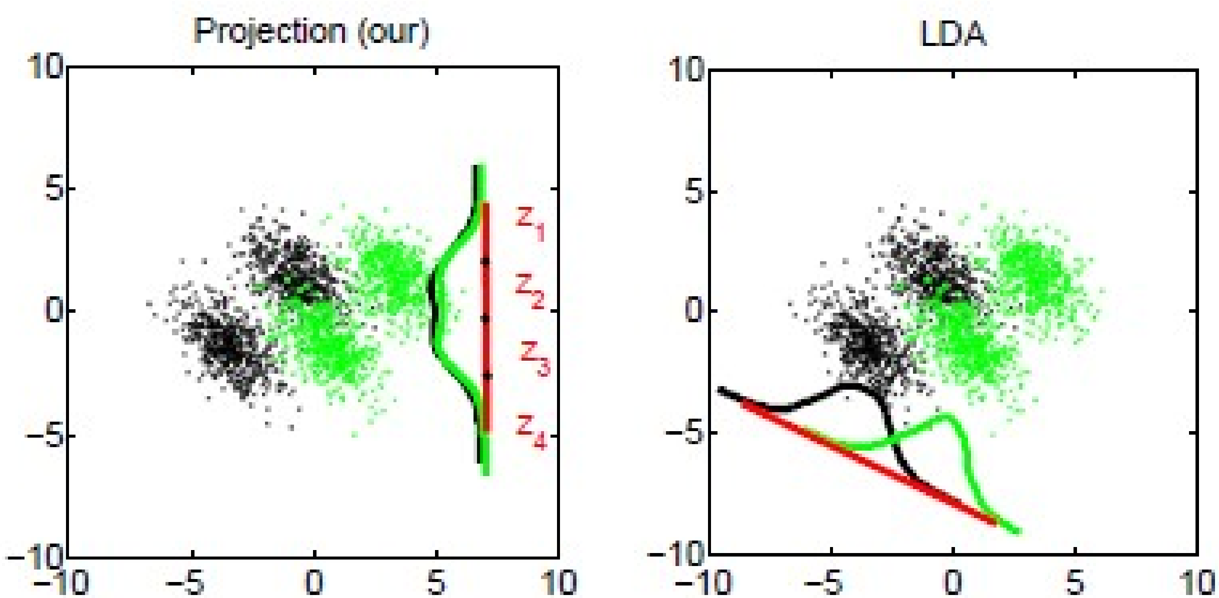
- Rotate the data $X' = V X$
- Maximizing the Fisher criterion $J^{LDA} = \frac{v^T S_b v}{v^T S_w v}$
  - maximizes between-class variance and minimizes with-class variance
- We want the opposite thus flip the criterion

$$J^{PR} = \frac{v^T S_w v}{v^T S_b v}$$

Projection

- Rotate the data $X' = V X$
- Maximizing the Fisher criterion $J^{LDA} = \frac{v^T S_b v}{v^T S_w v}$
  - maximizes between-class variance and minimizes with-class variance
- We want the opposite thus flip the criterion

$$J^{PR} = \frac{v^T S_w v}{v^T S_b v}$$



Feature underselection

- Measure correlations with individual features

$$r_i = corr(x_i, y)$$

- Discard the most correlated features

$$|r_{i1}| < |r_{i2}| < \ldots < |r_{ip-1}| < \cancel{|r_{ip}|}$$

- Cluster X'

Evaluation

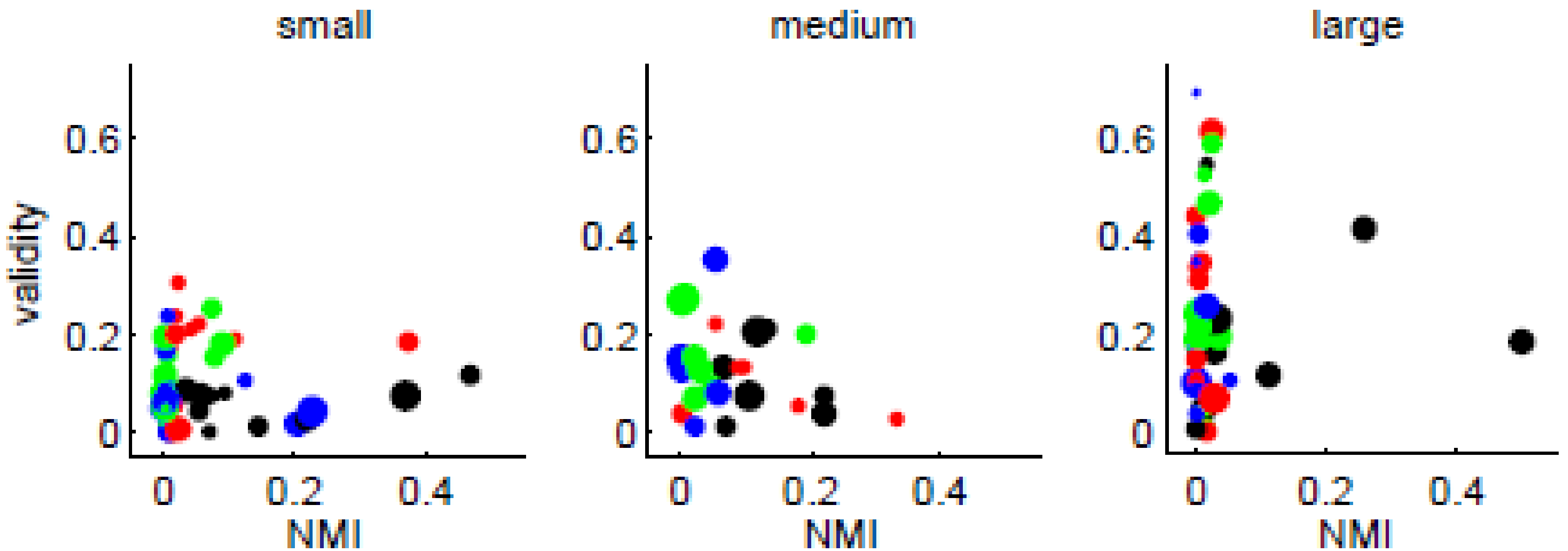
- How good are the identified contexts?
- Normalized mutual information for **independence** $NMI = \frac{I(y,z)}{\sqrt{H(y)H(z)}}$
- *But* random contexts would give $NMI = 0$
- Adapted measures from clustering
  - **Validity** learnable mapping from data to context $X \to z$
  - **Stability** $\{X_u \cup X_q\}$: train $clust_u(X_u)$ and $clust_q(X_q)$, $c_u = clust_u(X_q)$, $c_q = clust_q(X_q)$, compare c_u and c_q

Experimentals

- 30 classification datasets from UCI and other sources
- Size 500 – 67000, dimensionality 4-100, number of classes 2-11
- Split data into groups:
  - low-dimensional <10
  - medium-dimensional 10-19
  - high-dimensional >=20

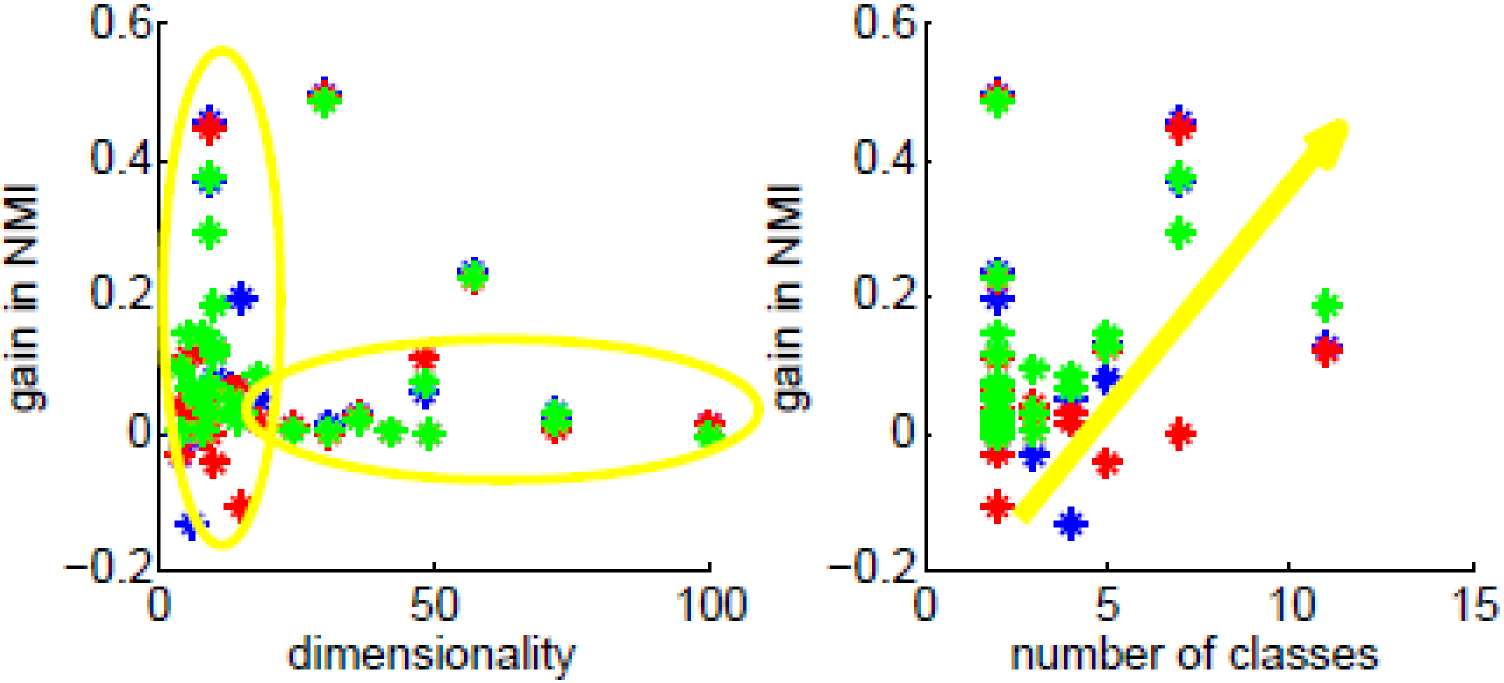
Results

- Transformation works well for low-medium dimensional data
- Less well, but ok for high dimensional data



Results

- Gain in independence w.r.t. simple clustering is larger for
  - low dimensional data and
  - data with larger number of classes



Case study

- Two level classification
  - transformation $X \rightarrow X'$
  - identify context $X' \rightarrow z$
  - select the model L_z

  Level 1

  - classify $L_z : X \rightarrow y$

  Level 2

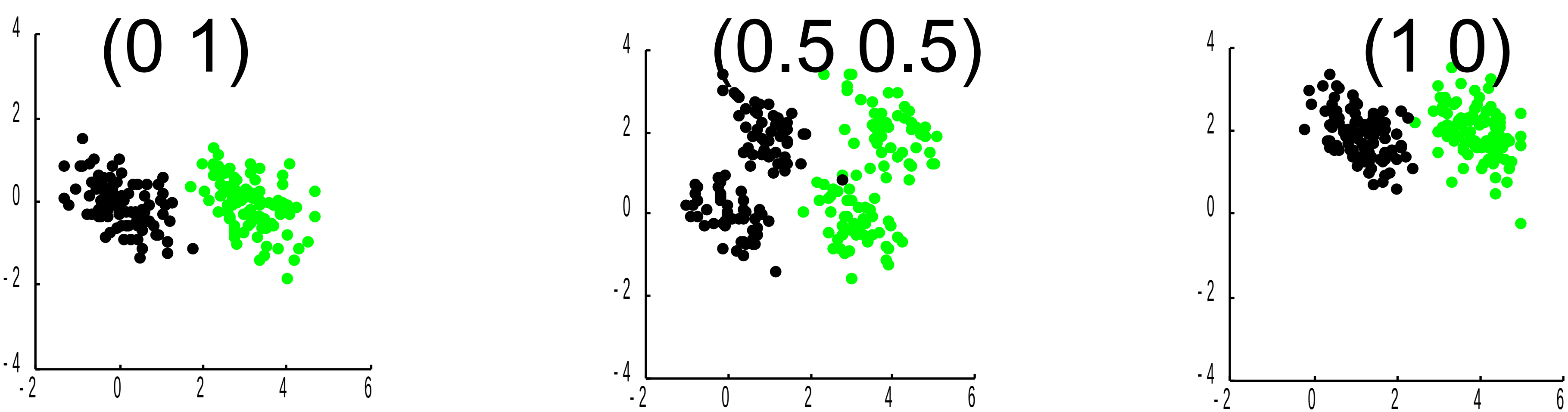
- Six base classifiers, 3 transformations
- K=2: EN ▶ PR ▶ FU ▶ OV ▶ AL ▶ CL ▶ RN
- K=4: EN ▶ OV ▶ FU ▶ CL ▶ PR ▶ AL ▶ RN
- K=7: EN ▶ PR ▶ AL ▶ OV ▶ FU ▶ RN ▶ CL

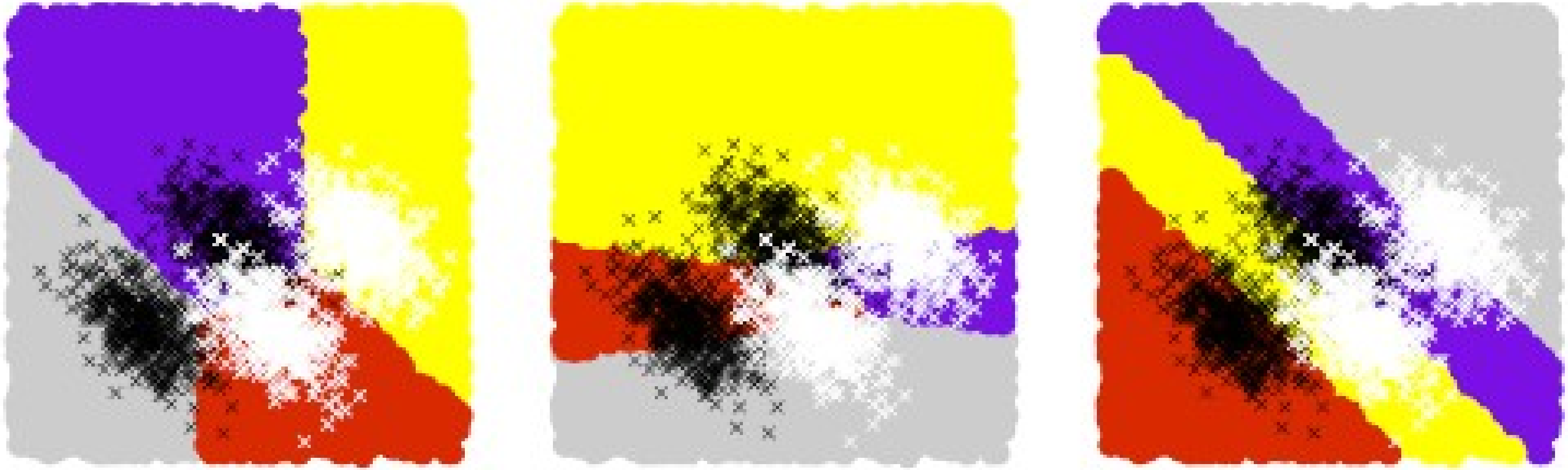
Outlook

- More trasfrmations for independence
  - Clustering with constraints
  - Decision trees with 'twisted' splitting criteria
- Related tasks: managing hidden dependencies
- Modeling evolving data as a **mixture** of contexts

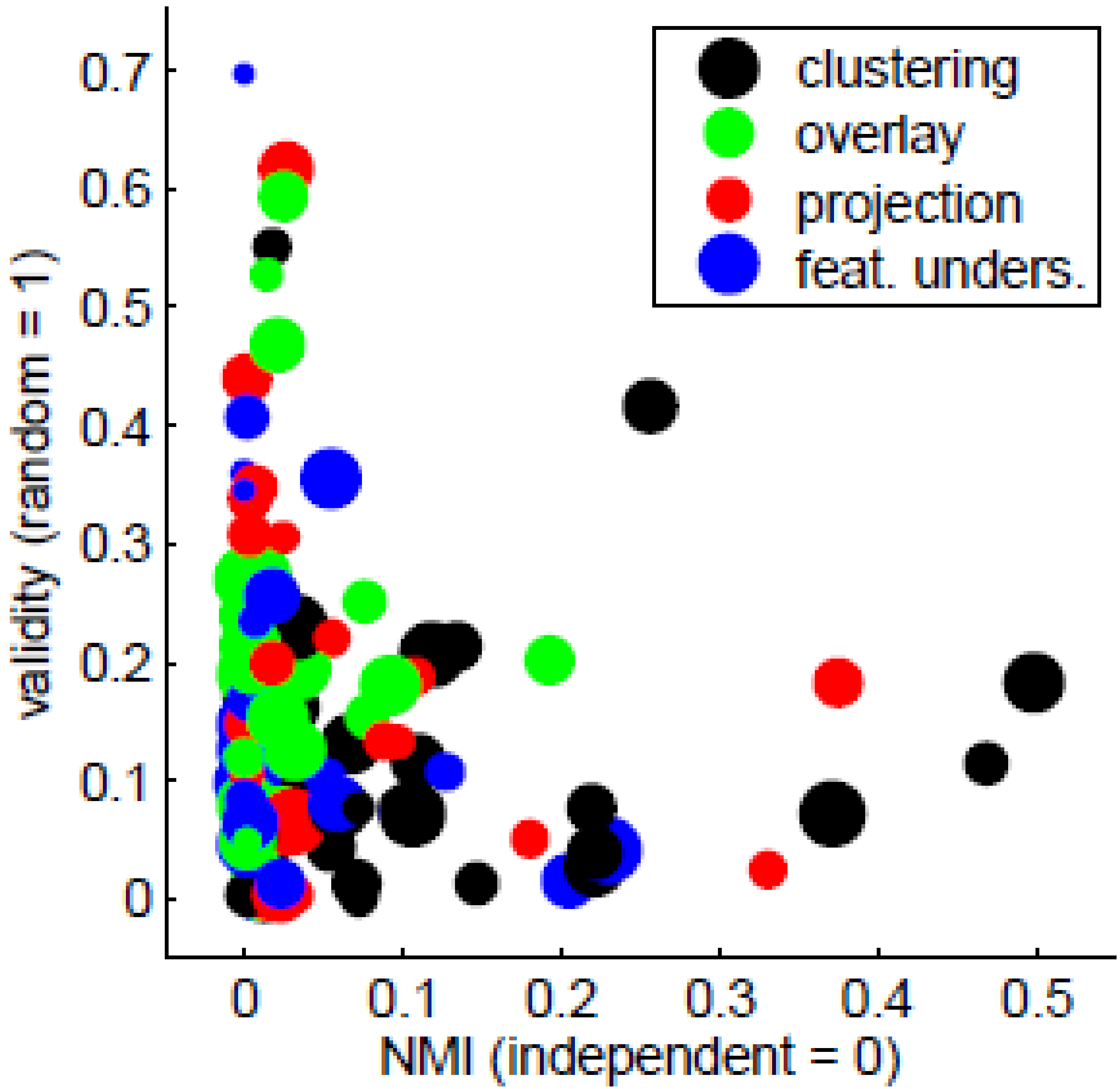
Outlook

- More trasfrmations for independence
  - Clustering with constraints
  - Decision trees with 'twisted' splitting criteria
- Related tasks: managing hidden dependencies
- Modeling evolving data as a **mixture** of contexts

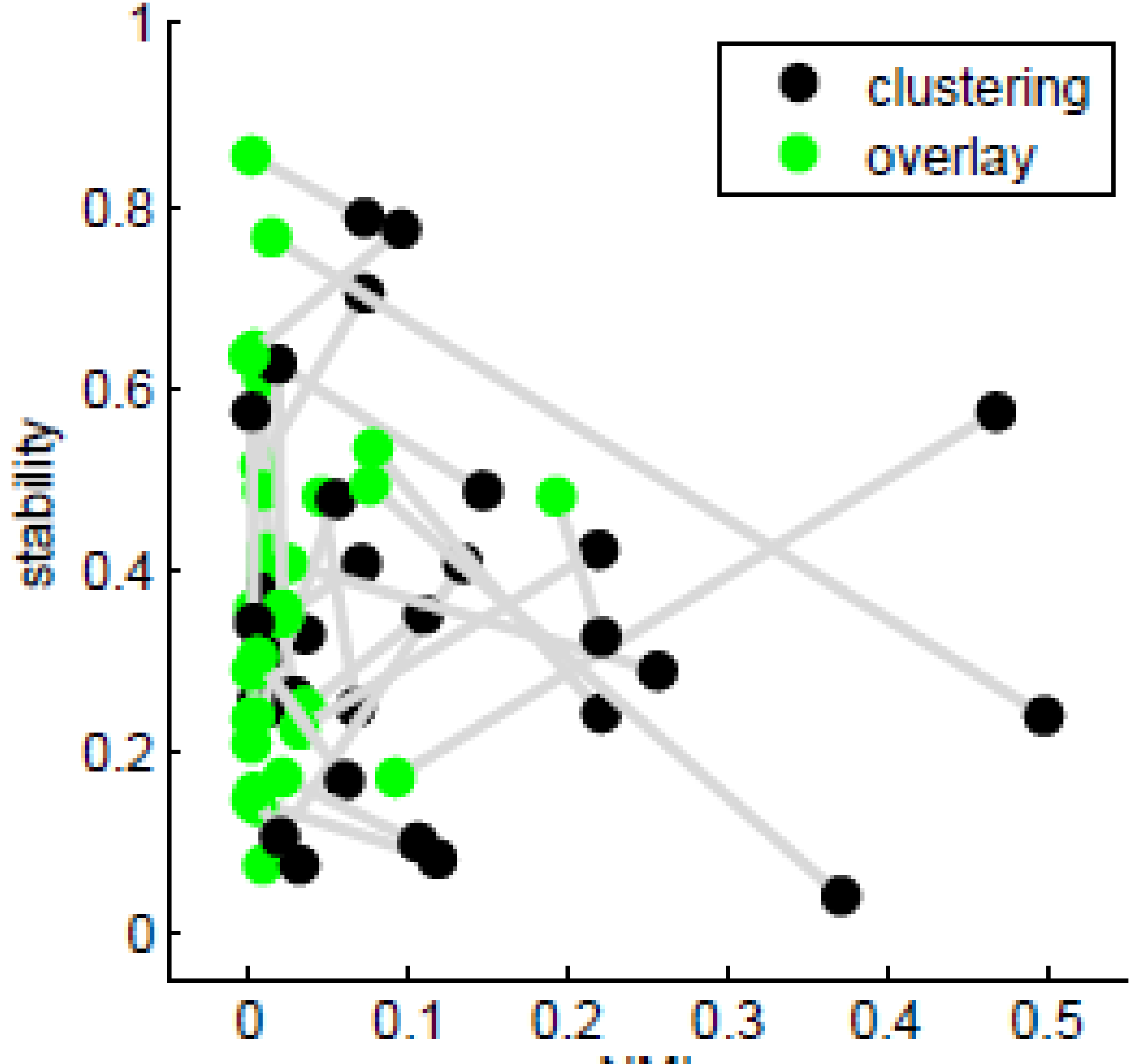




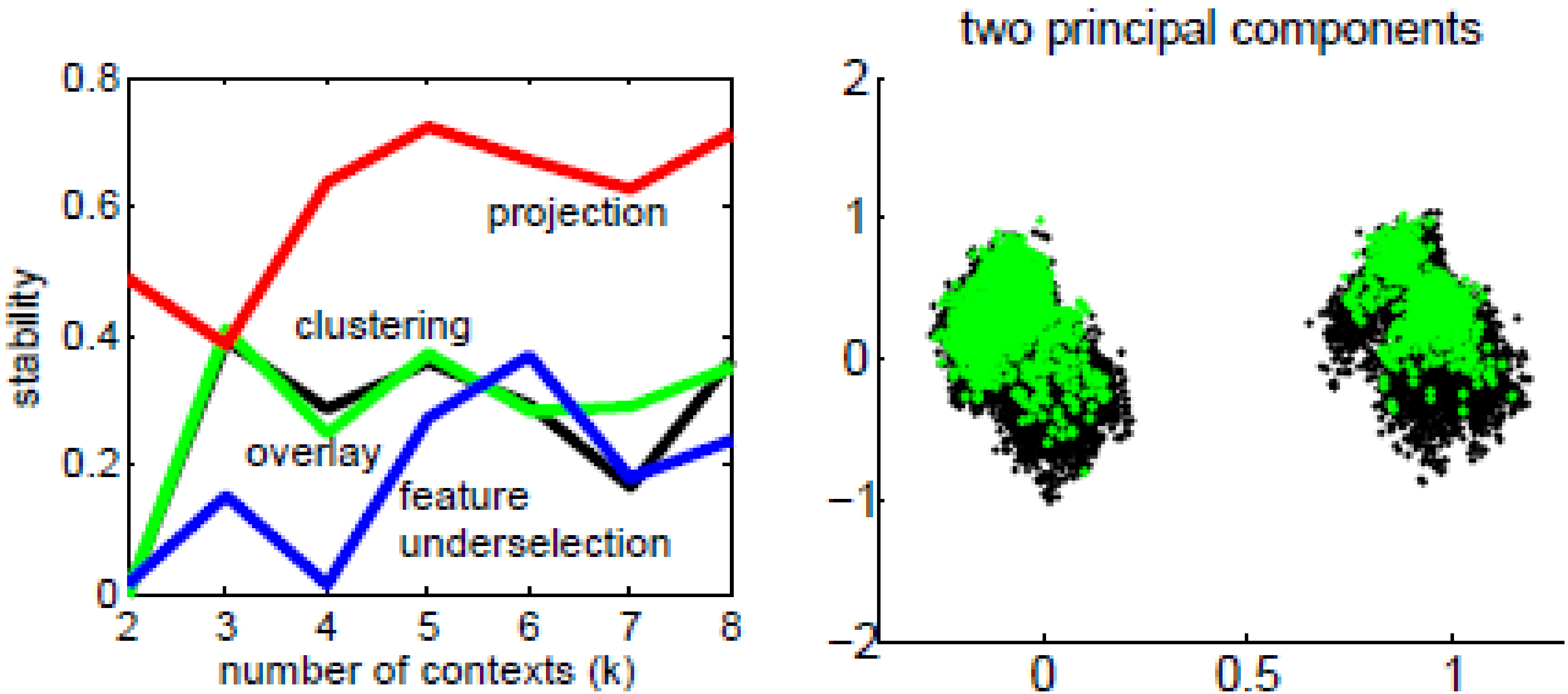
(a) clustering (b) overlay (c) projection



clustering
overlay
projection
feat. unders.
validity (random = 1)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
NMI (independent = 0)
0
0.1
0.2
0.3
0.4
0.5



clustering
overlay
stability
NMI
0
0.2
0.4
0.6
0.8
1
0
0.1
0.2
0.3
0.4
0.5



stability
0.8
0.6
0.4
0.2
0
2
3
4
5
6
7
8
number of contexts (k)
projection
clustering
overlay
feature
underselection
two principal components
2
1
0
−1
−2
0
0.5
1