UNIVERSITY OF TWENTE.

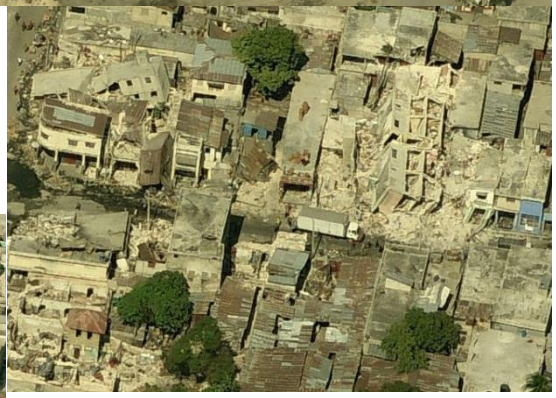
Role of dimensionality reduction in segment-based classification of damaged building roofs in airborne laser scanning data

Kourosh Khoshelham

ITC FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Detection of damaged buildings in post-disaster aerial data

- **Why detect damaged buildings?**
  - → Planning for the recovery phase;
  - → Rebuilding damaged buildings;
  - → Repairing infrastructure.
- **Why automatically?**
  - → Manual procedure: too slow/expensive.



Detection of damaged buildings in post-disaster aerial data

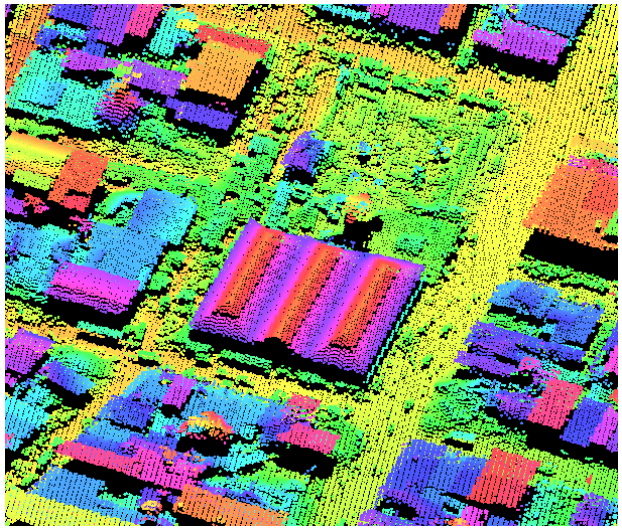
Aerial images:

→ Convenient for human interpretation

Laser scanning point clouds:

→ More accurate representation of geometry;

→ More accurate classification.



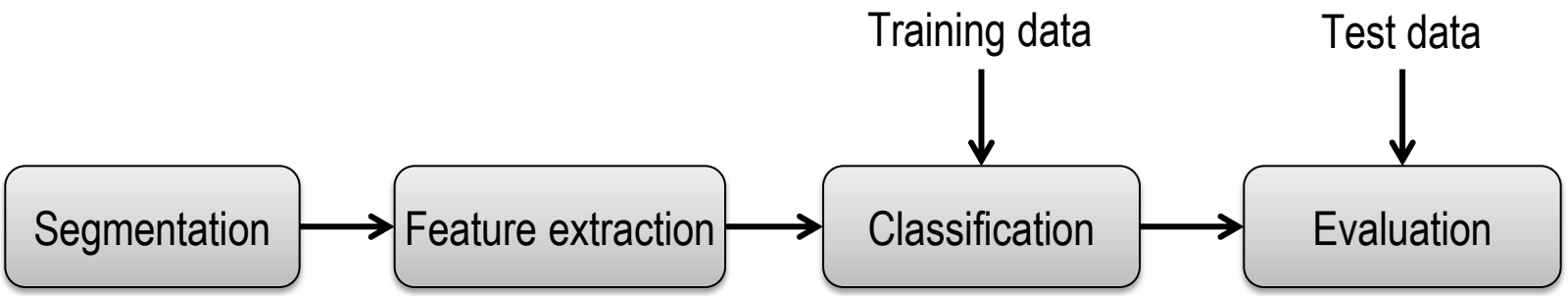
Segment-based classification of point clouds

- **Assumption:**

  → Intact building roofs comprise a few large planar segments;

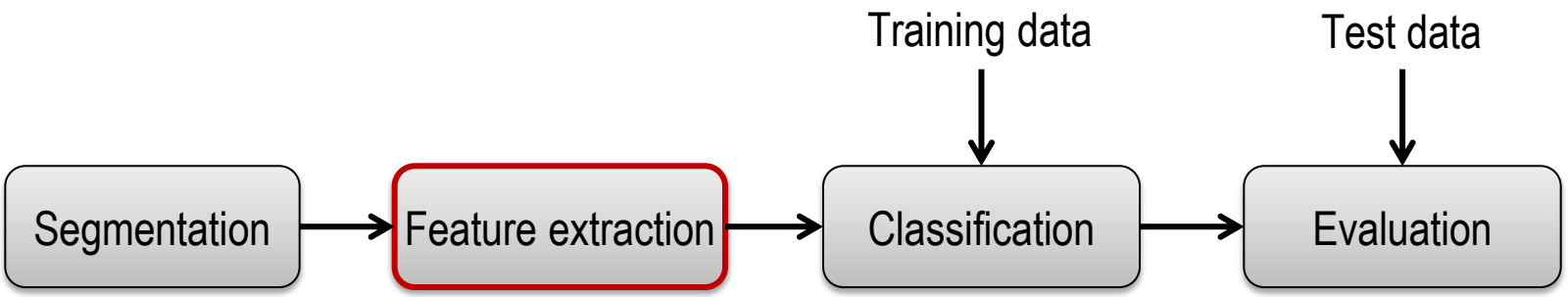
  → Damaged roofs appear as many small segments.

- Possibly relevant features: segment size, planarity, orientation, height, …?



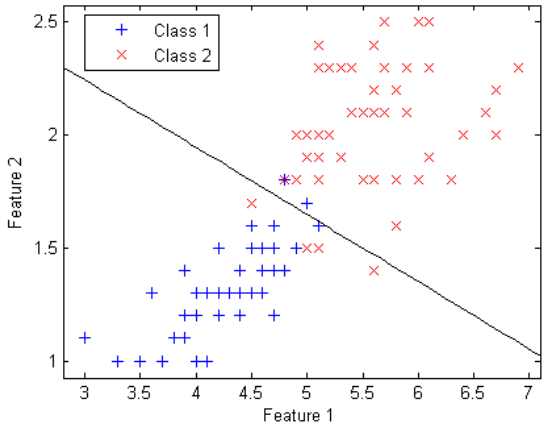
Segment-based classification of point clouds

- **Dimensionality problem:**
  - → Lack of knowledge about relevant features;
  - → Include many features?
  - → More features = more training samples;
  - → Training samples insufficient due to difficulty of interpreting point clouds;
  - → Complex classifiers?

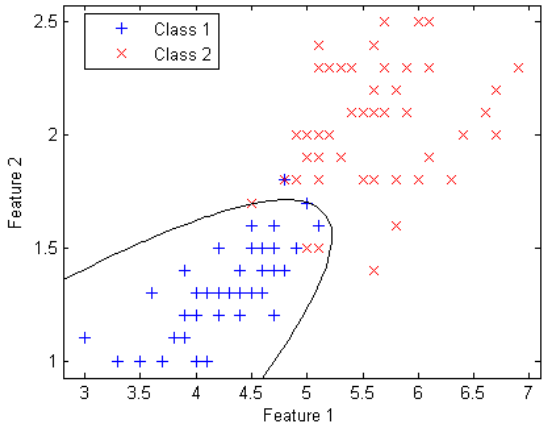


Classifier complexity

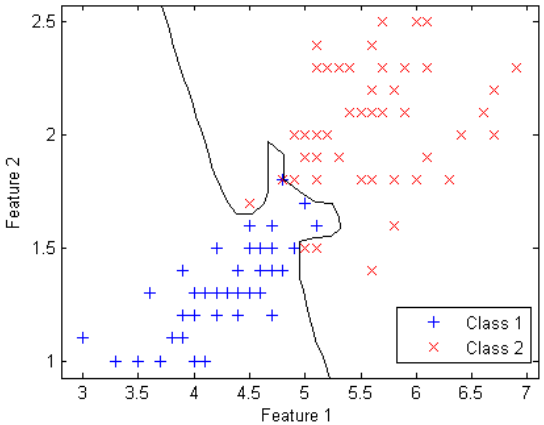
Linear discriminant classifier



Quadratic discriminant classifier

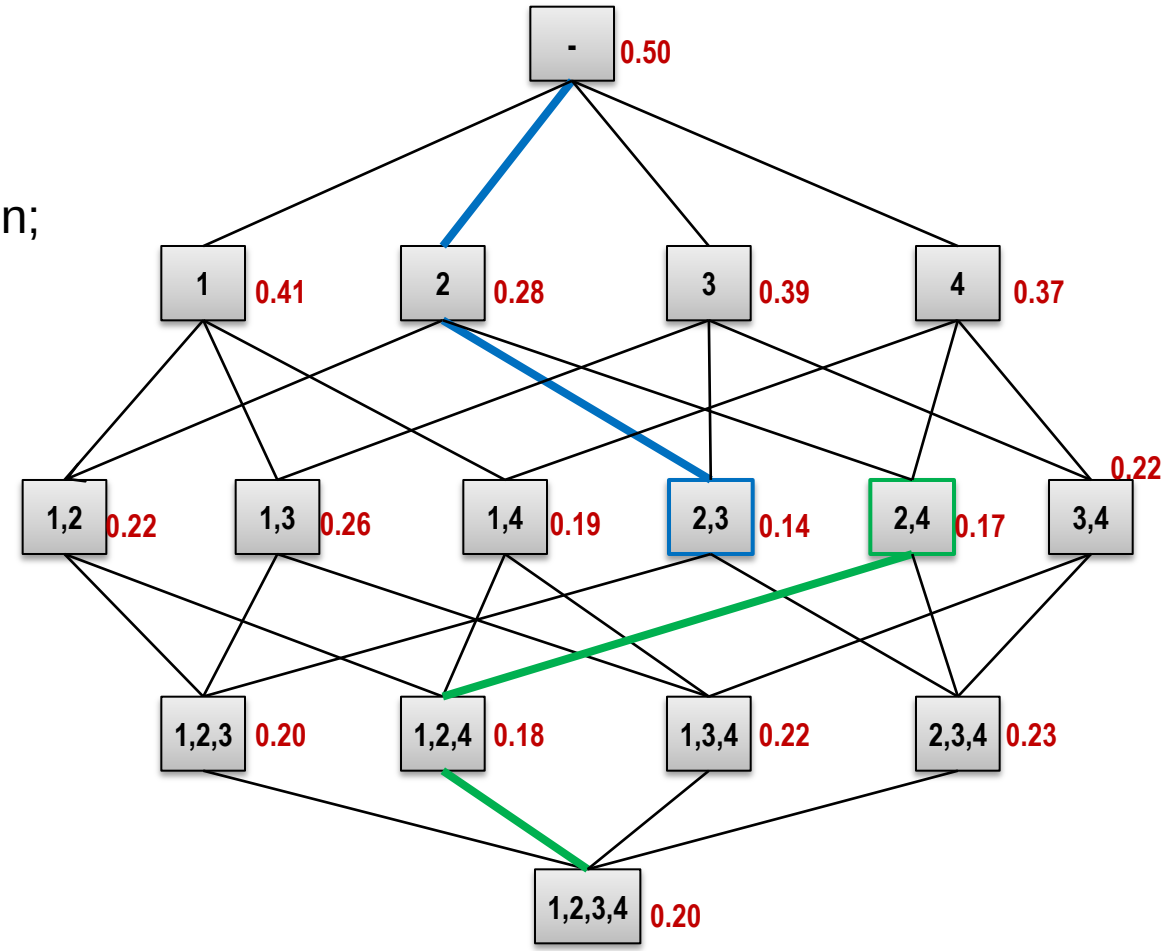


Nearest neighbor classifier



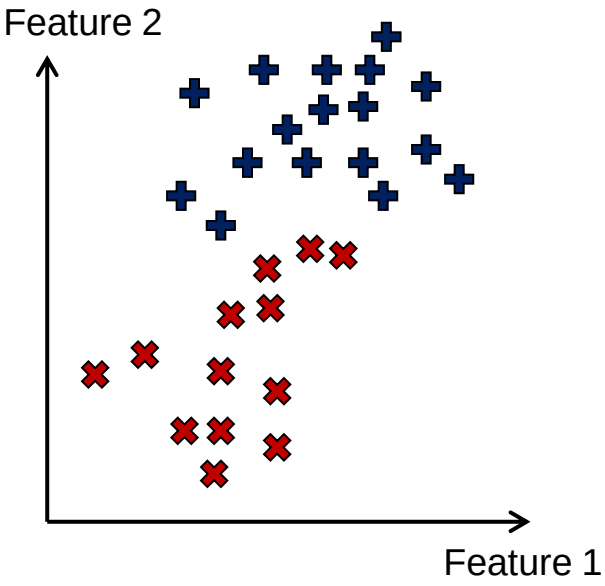
Dimensionality reduction

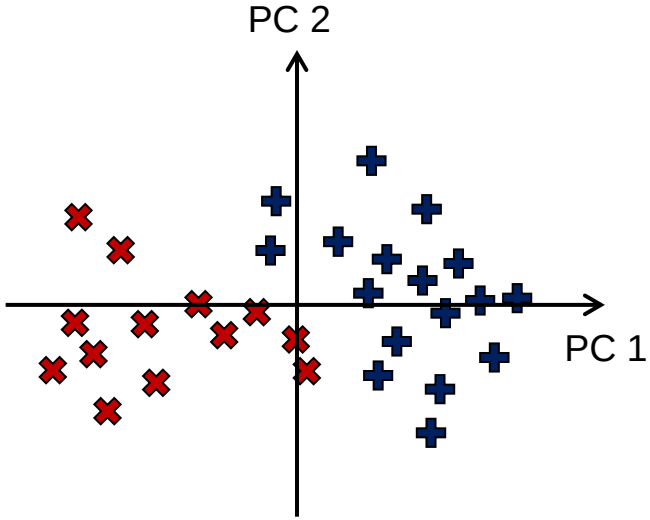
- **Feature selection:**
  - → Forward selection;
  - → Backward elimination;
  - → Plus *l* take away *r*.
  - → Branch and bound.

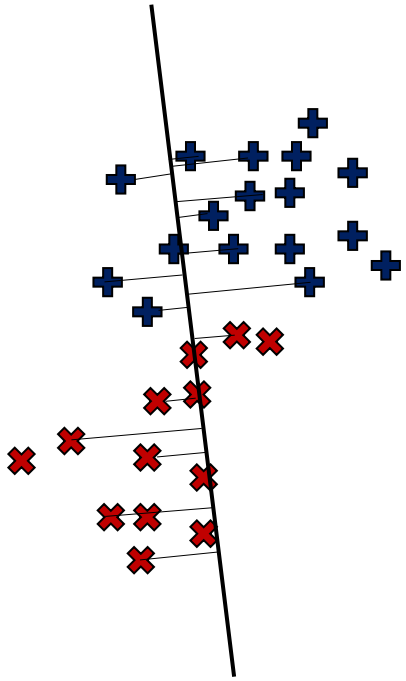


Dimensionality reduction

- **Mapping features to reduced dimensions:**
  - → Principal Component Analysis;
  - → Linear Discriminant Analysis.

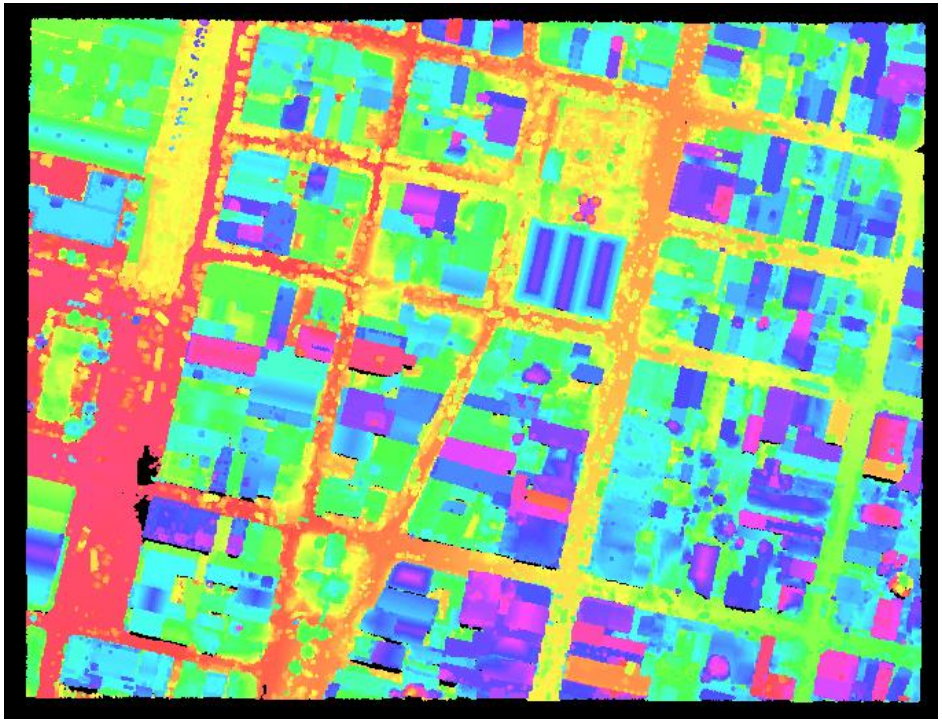






LDA mapping

Experiments



Airborne laser data (3 pnts/m2) of Port-au-Prince after the earthquake of Jan. 2010.

feat. id	Description
1	number of points per segment (nP)
2	root mean square of plane fitting residuals (rmsRes)
3	ratio of plane fitting outliers (rOut)
4	plane slope (s)
5	z component of plane normal (n_z)
6	mean reflectance per segment (meanRfl)
7	standard deviation of reflectance per segment (stdRfl)
8	minimum height above DTM (minH)
9	maximum height above DTM (maxH)
10	mean height above DTM (meanH)
11	ratio of points in a segment that have an unsegmented point in a neighbourhood of 1 m (rUnseg)
12	mean point density in the bounding box of the segment (meanDns)

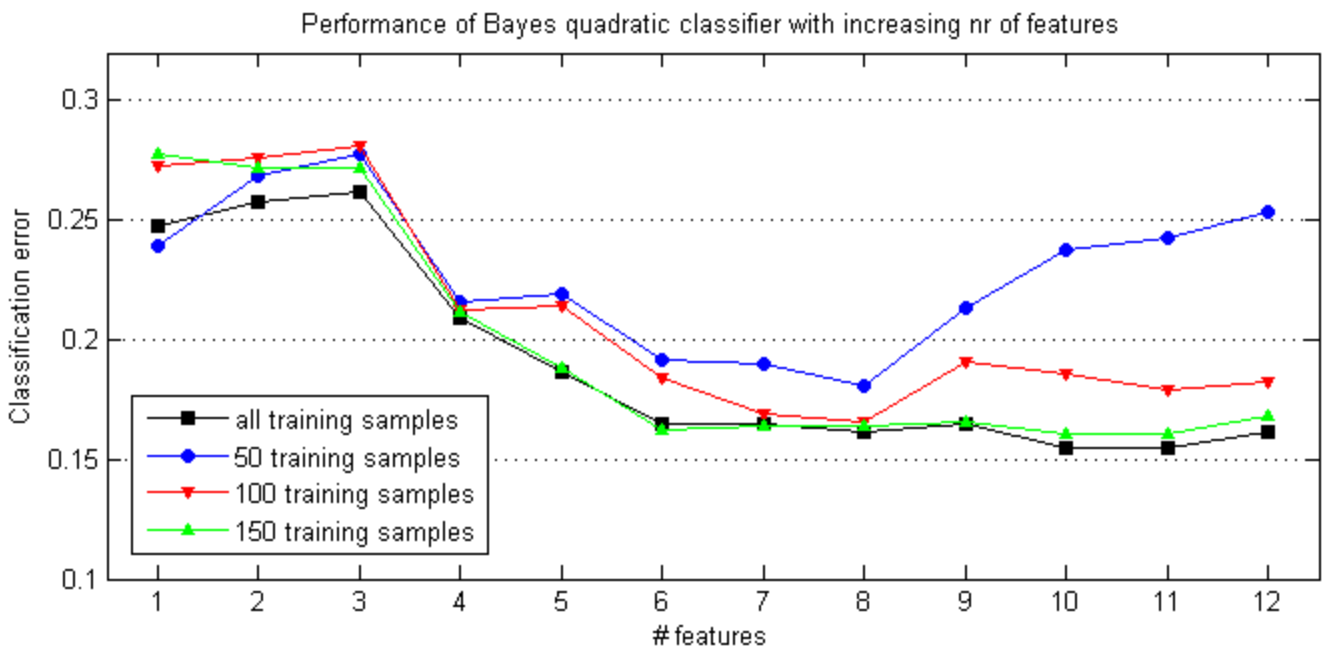
Features extracted per segment

Results

Number of features Vs number of training samples

→ With >50 training samples no contribution from last 6 features;

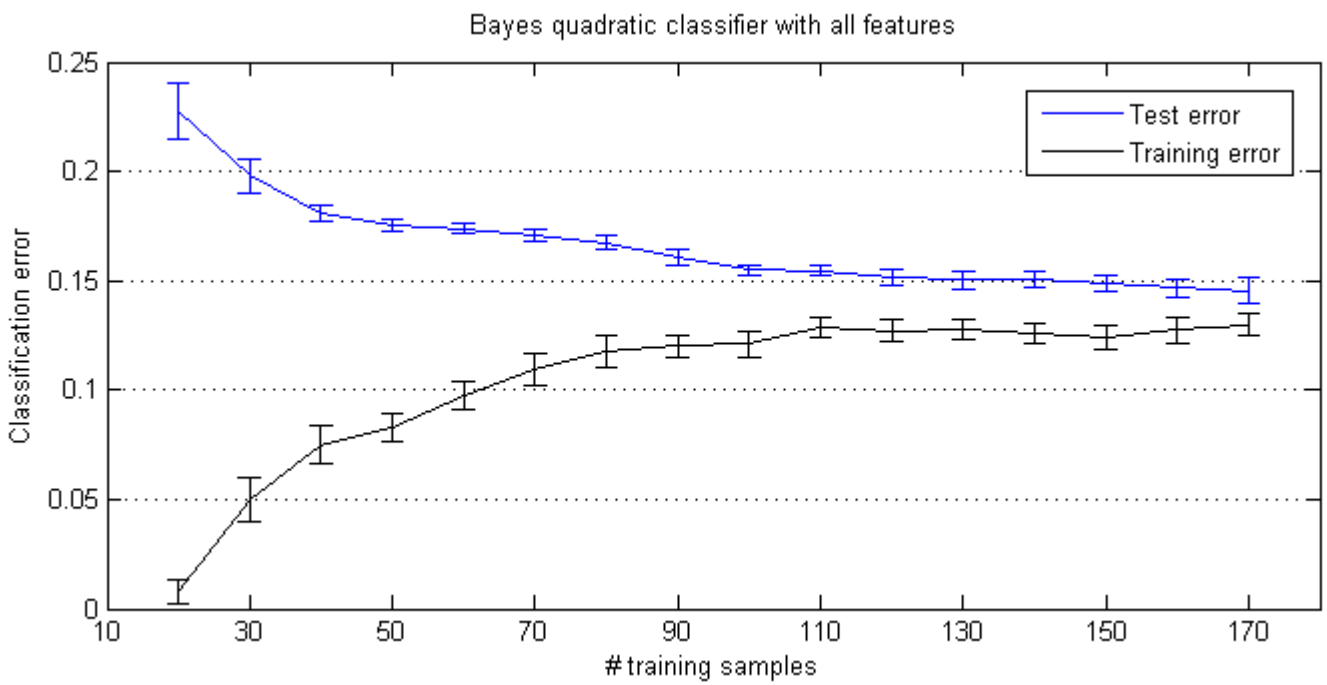
→ With 50 training samples weaker features deteriorate classifier's performance.



Results

Number of training samples

→ Minimum 100 samples needed to avoid large bias (difference between test error and training error).

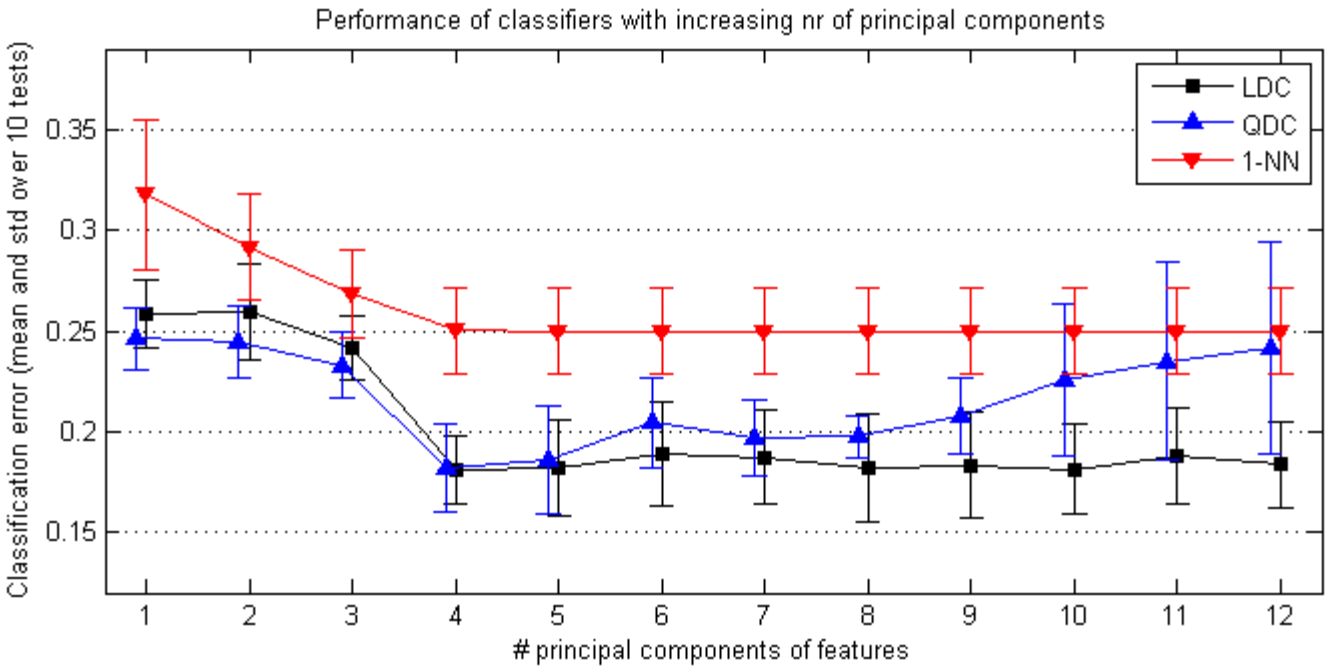


Results

Classification with principal components of features

→ Only the first 4 principal components are useful;

→ Linear classifier outperforms more complex ones.



UNIVERSITY OF TWENTE.

Results

Feature selection

→ No global optimal subset;

→ More selected features:

- → 1: number of points per segment;
- → 3: ratio of plane fitting outliers;
- → 10: mean height above DTM;
- → 11: ratio of points located near an unsegmented point.

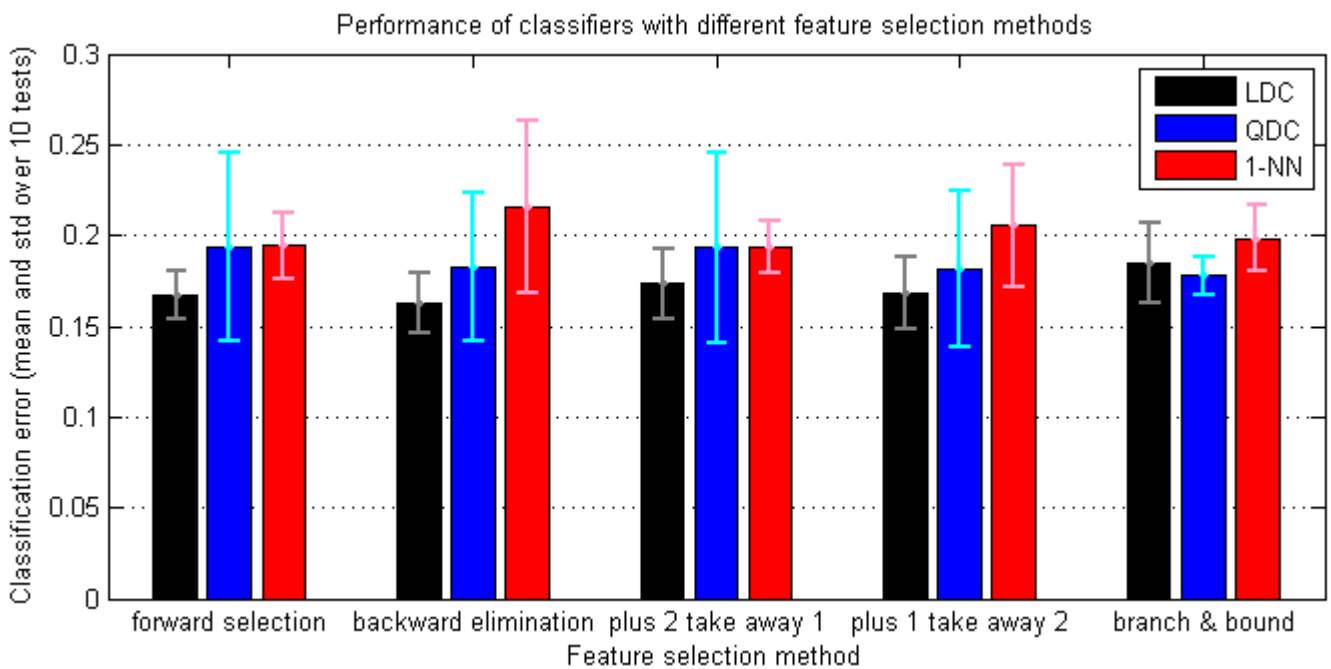
Search method	Classifier	Error	Features											
			1	2	3	4	5	6	7	8	9	10	11	12
FS	LDC	0.167	X									X	X	X
	QDC	0.194	X	X	X	X	X	X	X			X	X	
	1-NN	0.195		X	X	X	X					X	X	
BE	**LDC**	**0.163**	**X**		**X**	**X**			**X**	**X**		**X**	**X**	
	QDC	0.183	X	X	X		X		X			X	X	
	1-NN	0.216		X	X					X		X		X
+2-1	LDC	0.173	X		X	X	X		X	X	X	X	X	
	QDC	0.193	X	X	X	X	X	X	X	X			X	
	1-NN	0.194		X		X						X	X	
+1-2	LDC	0.168	X		X		X		X		X	X	X	
	QDC	0.181	X	X	X		X		X	X			X	
	1-NN	0.206		X	X					X	X	X	X	
BB	LDC	0.185	X									X		
	QDC	0.178										X	X	
	1-NN	0.198									X	X		

Results

Comparison of feature selection methods

→ Similar performance of feature selection methods;

→ Linear classifier performs better than more complex ones.



UNIVERSITY OF TWENTE.

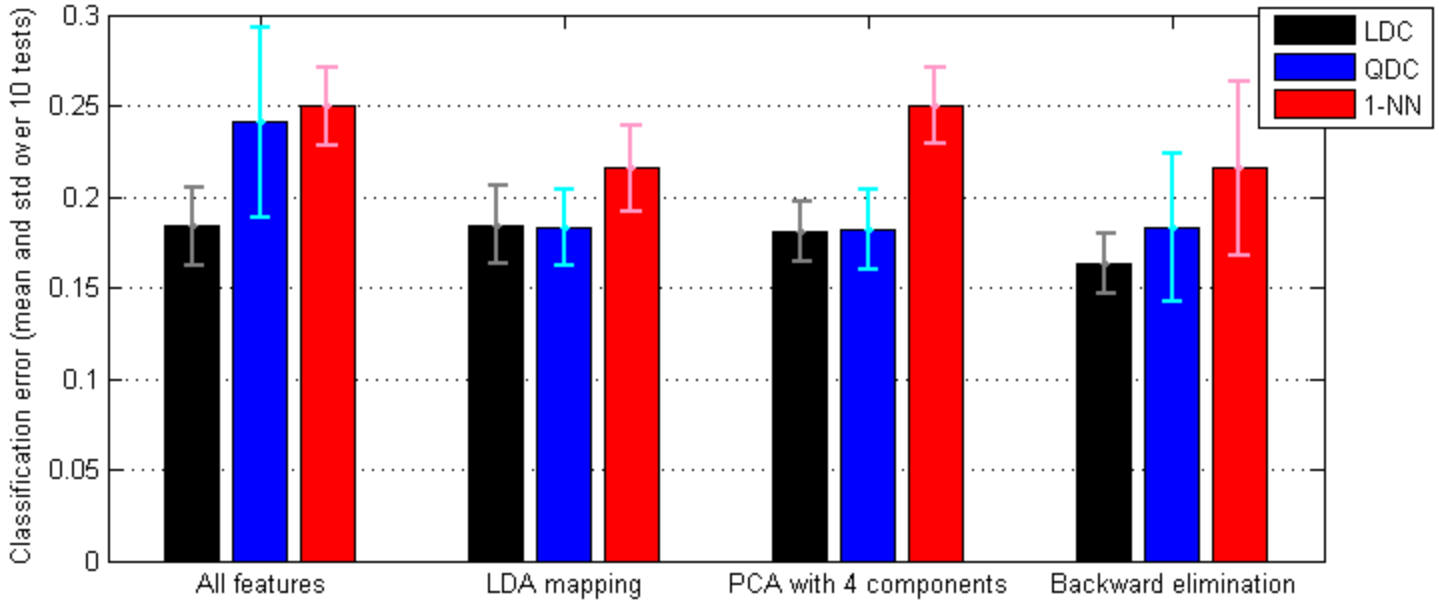
Results

Feature selection Vs mapping to lower dimensions Vs complexity

→ Dimensionality reduction in general improves classification results;

→ LDC+BE = 84% classification accuracy

→ LDC best, 1-NN worst, QDC improves more with dimensionality reduction.

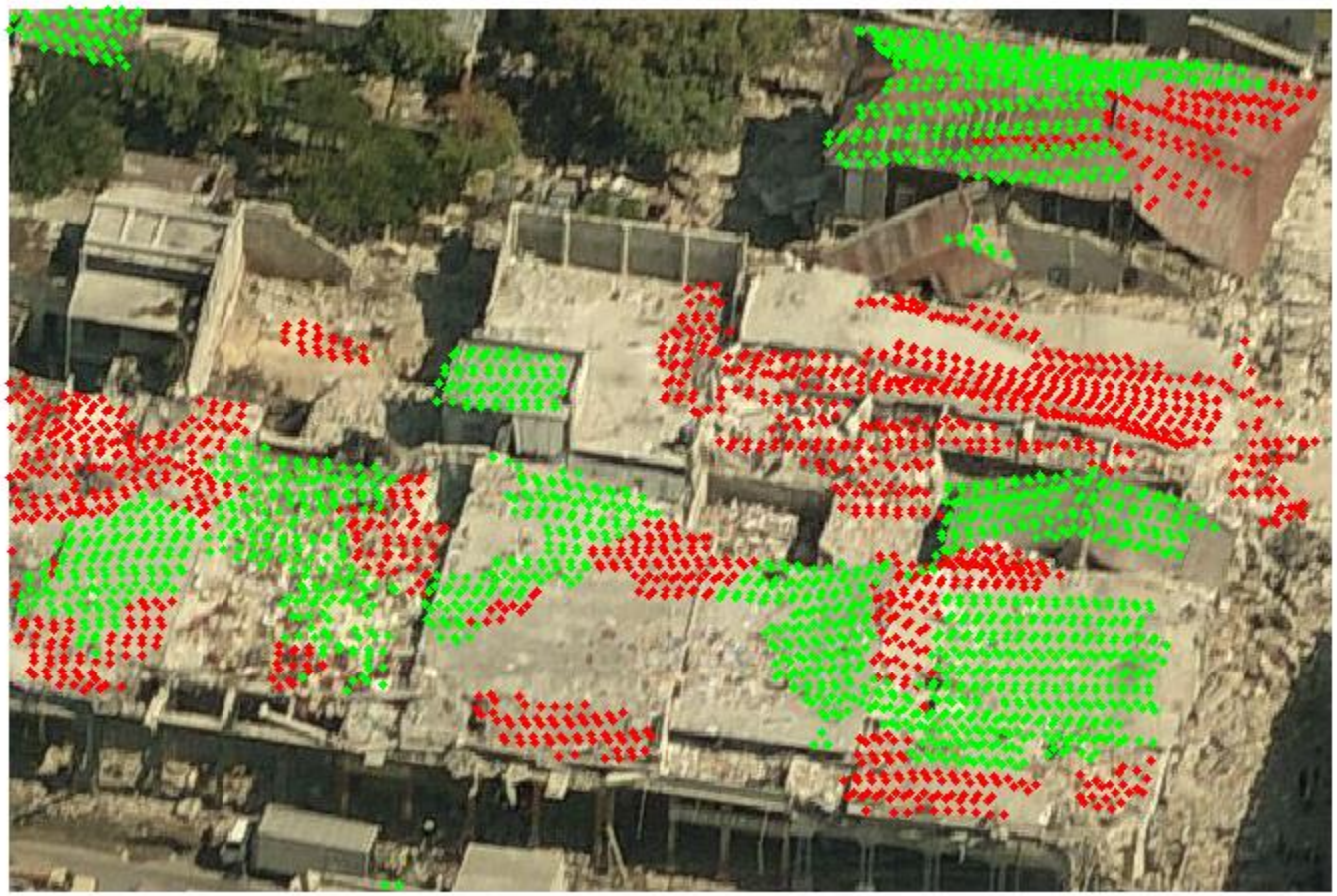


Results

Results

Results

Results



Summary and concluding remarks

- **Segment-based classification of laser data**
  - → Relevance of features;
  - → Training data collection;
  - → More features = more training samples;
- **Dimensionality reduction**
  - → Both feature selection and mapping to reduced dimensions improve classifier performance;
- **Classifier complexity**
  - → Less complex classifiers perform better when the number of training samples in proportion to the number of features is small.
- **Future work:**
  - → Considering more classes (too avoid confusion between them).