MCS 2005

Combining Feature Subsets in Feature Selection

Marina Skurichina, Robert P.W. Duin
June 2005

Delft University of Technology
The Netherlands

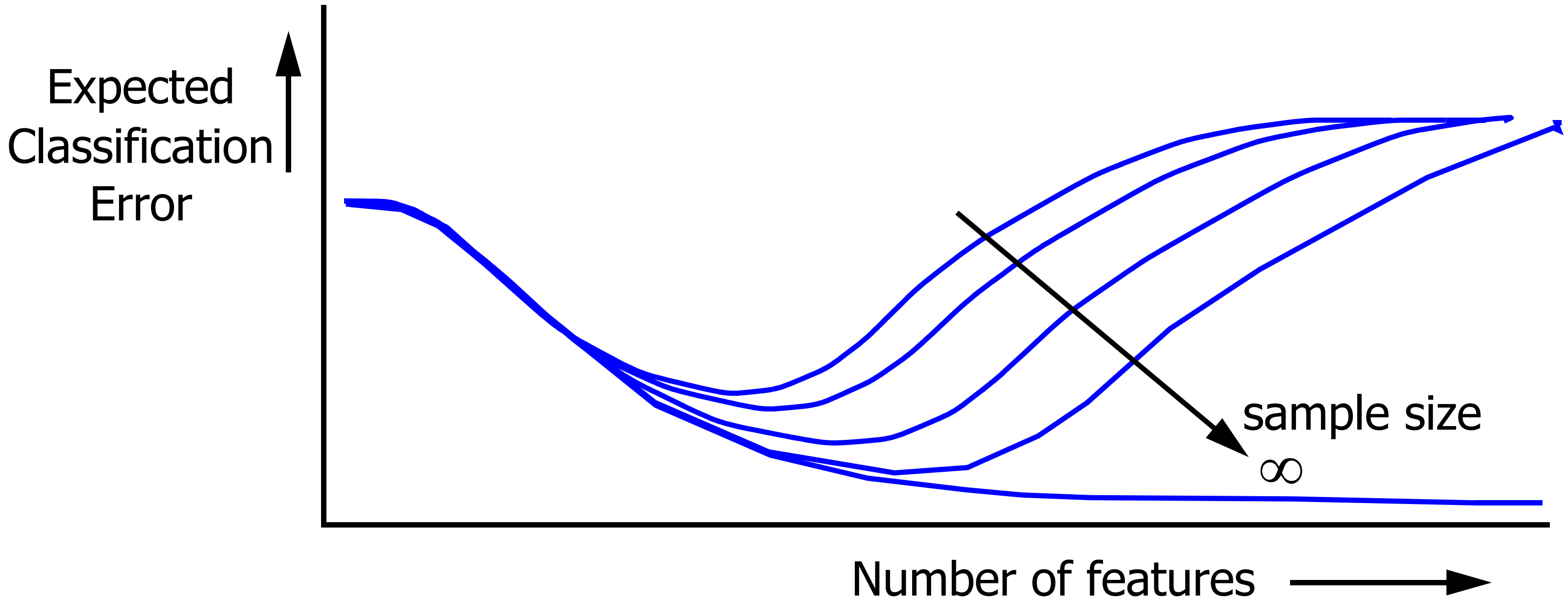
TU Delft

Feature Selection

Reasons for feature selection:

1. Cost reduction: measurements, computations

2. Accuracy



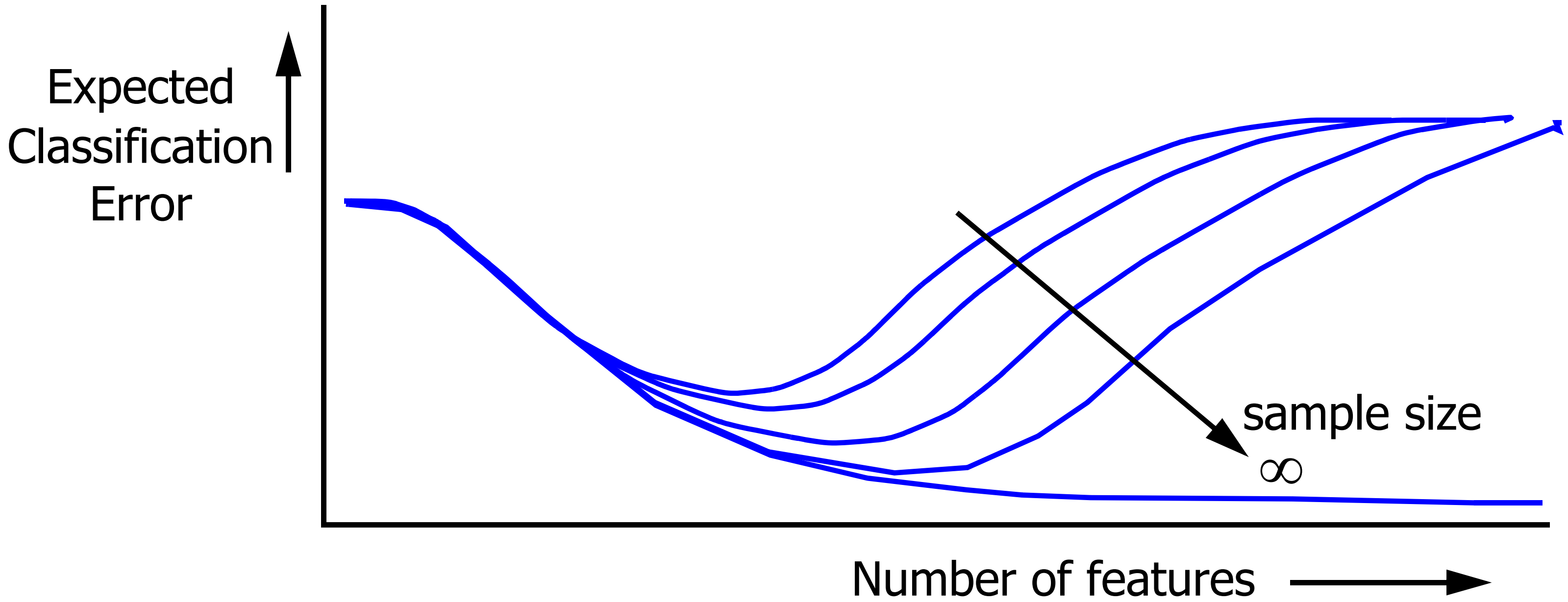
TU Delft

Feature Selection

Reasons for feature selection:

1. Cost reduction: measurements, computations

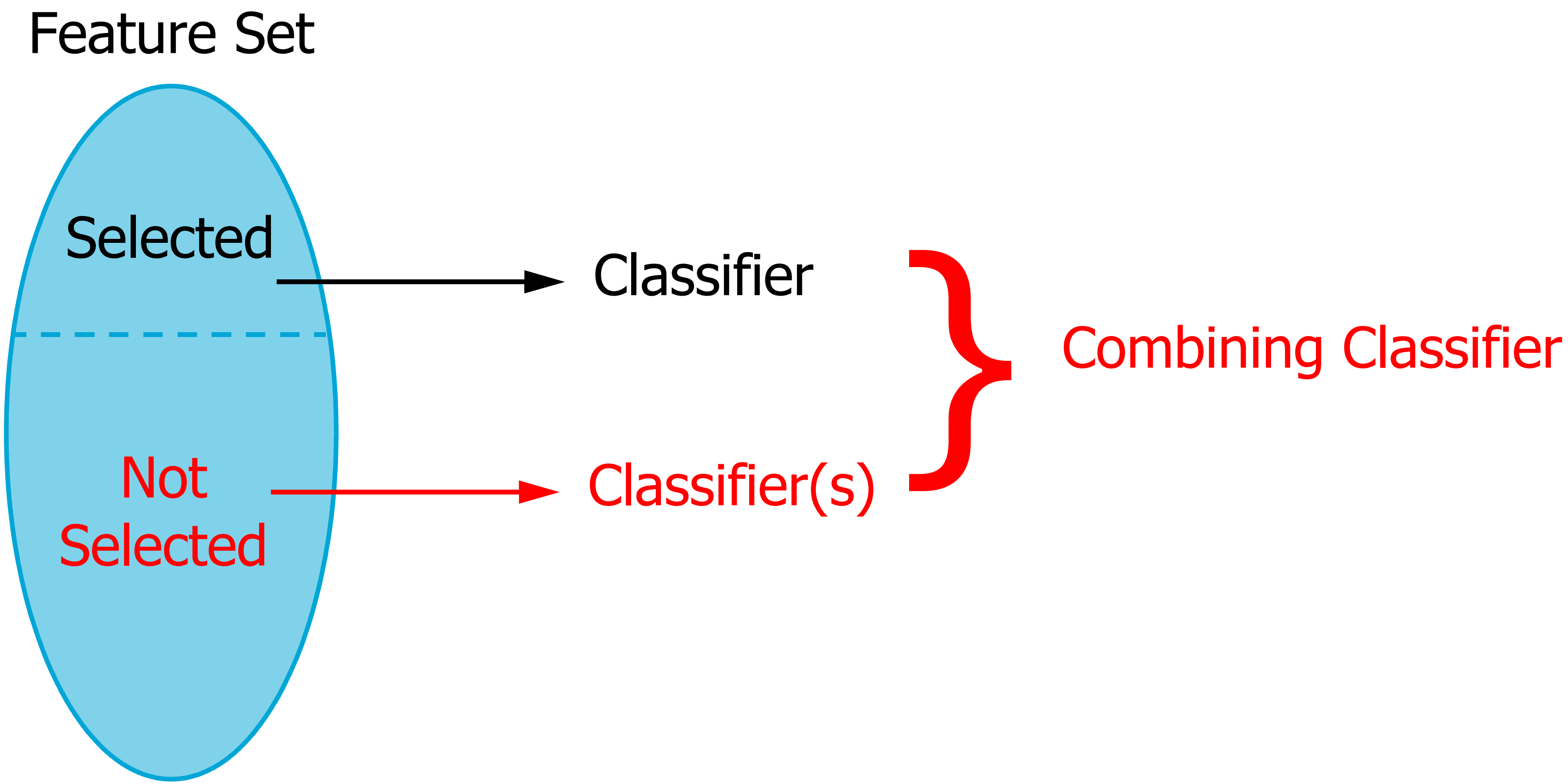
2. Accuracy ?



TU Delft

Question 1

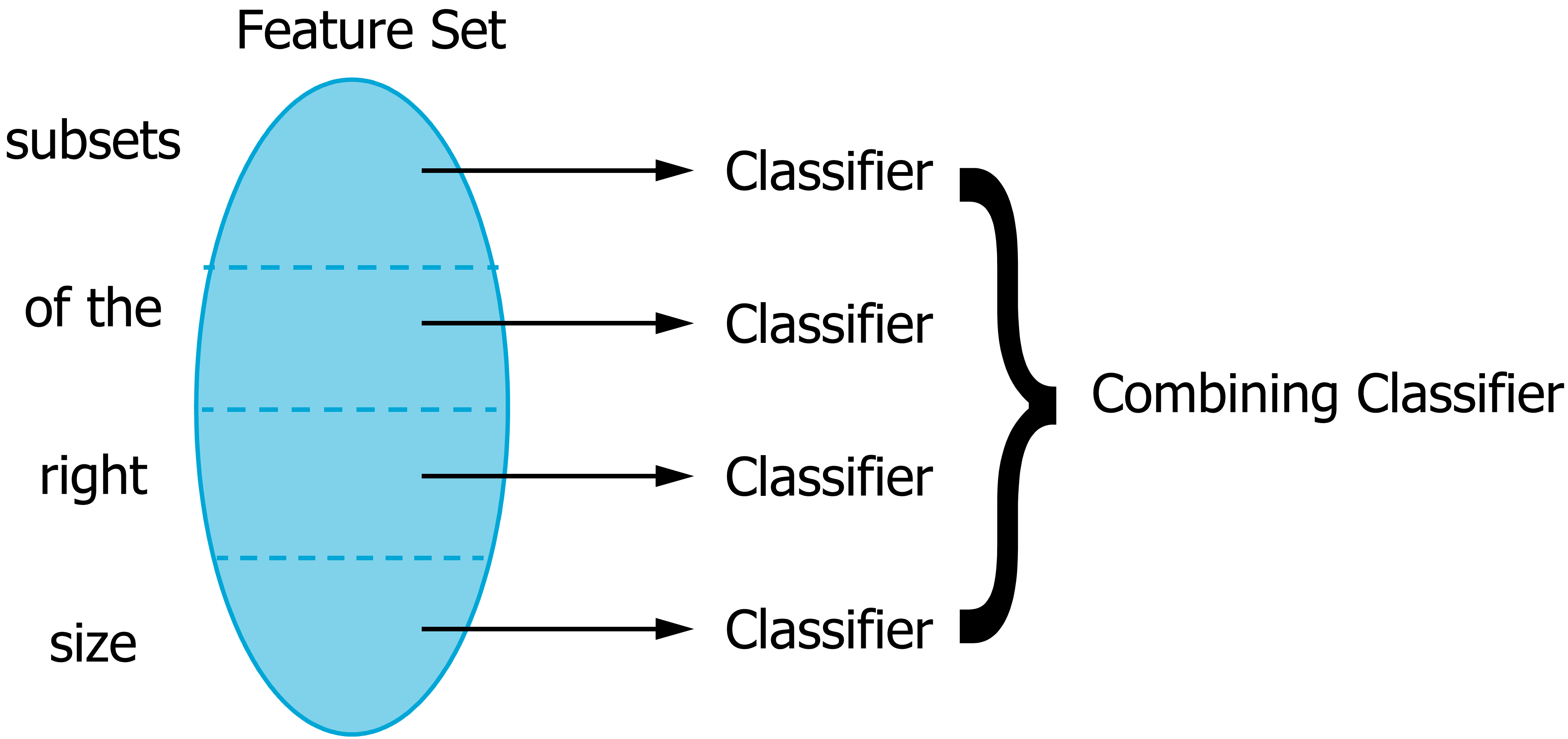
Can not-selected features be used to improve classifiers based on the selected features?



TU Delft

Question 2

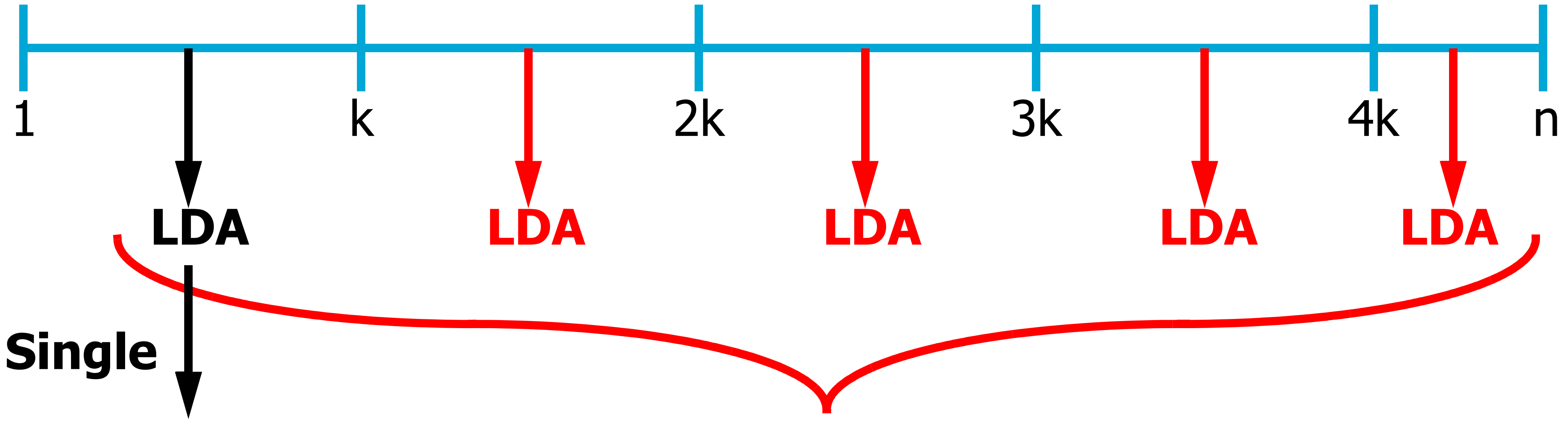
Is feature selection needed anyway?



TU Delft

Feature Selection and Ranking

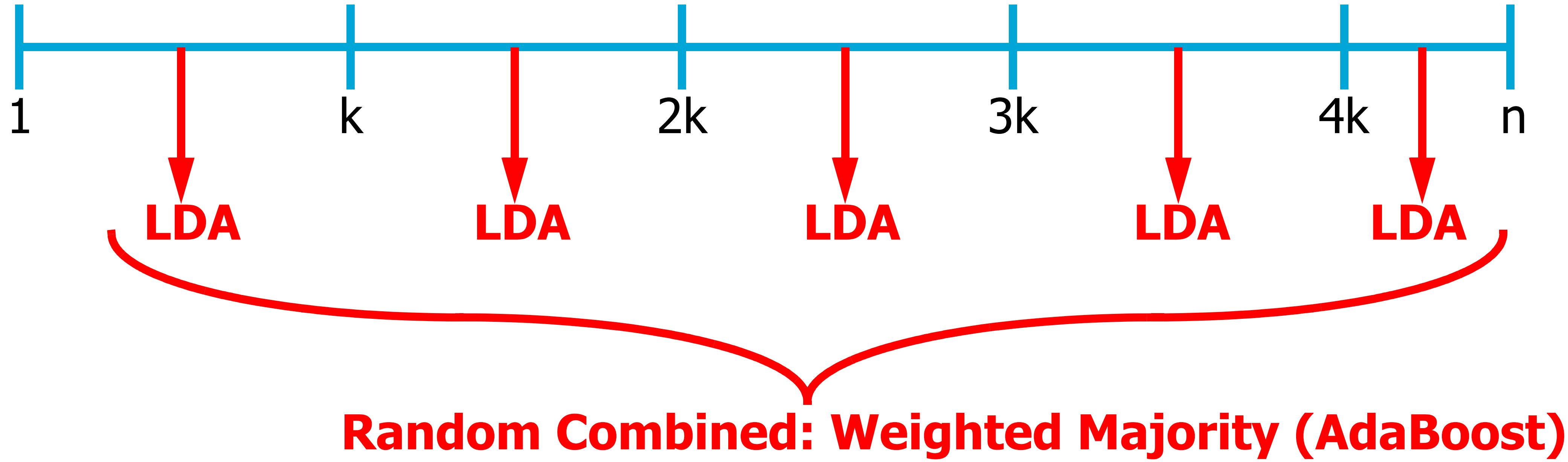
Rank features by forward selection - 1-NN Leave-one-out performance



Forward-Combined: Weighted Majority (AdaBoost)
Mean
DT (Decision Template)

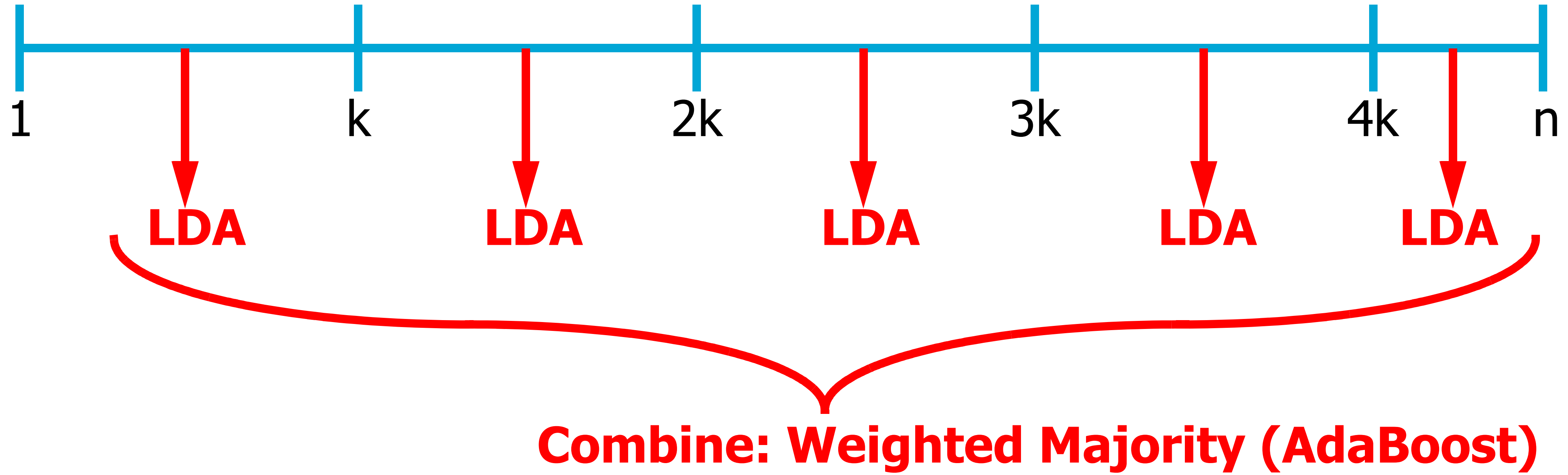
Random Feature Ranking

Rank features randomly:



Random Subspace Method

Select repeatedly k out of n features, in total n: doubles and missing!



Experiments

1. Take a dataset.

2. Split at random in train and test set.

3. Rank features.

   - forward selection.

   - random selection (no duplicates, no missings).

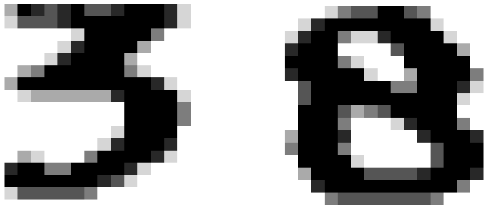
   - random subspace (duplicates, missings).

4. Compute for all feature subset sizes (combined) LDA classifiers.

5. Evaluate by test set.

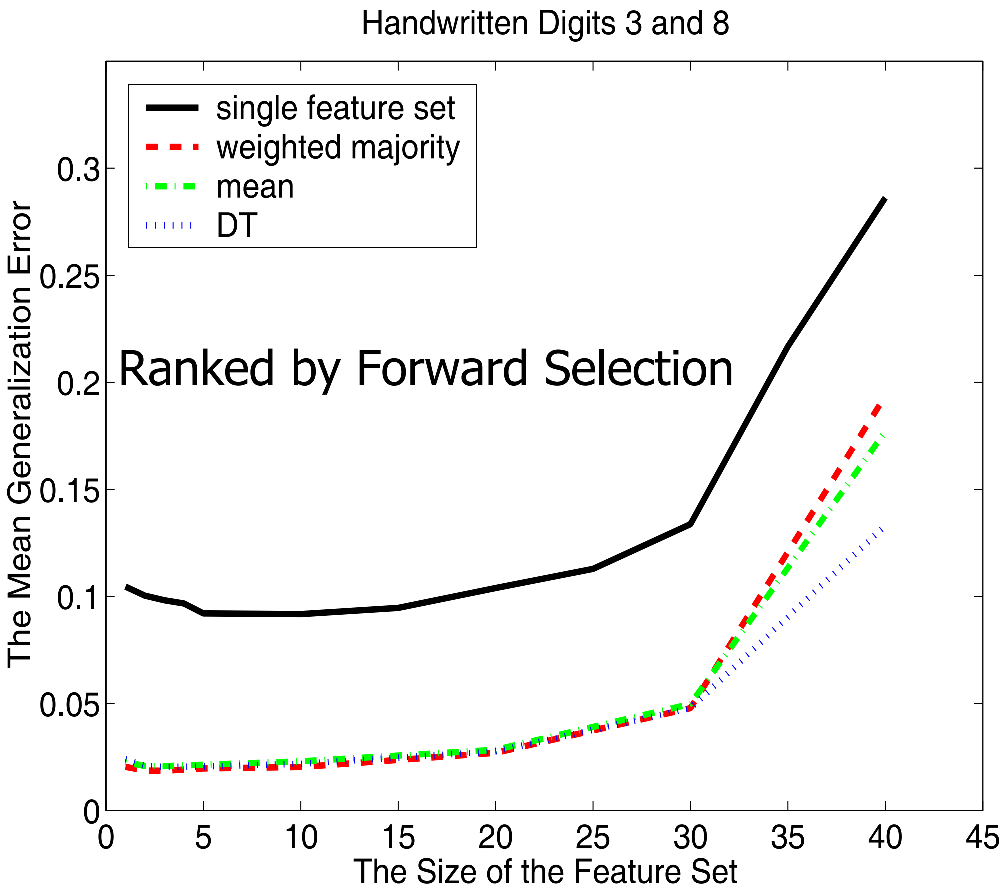
6. Repeat 2-5 50 times, average error estimates.

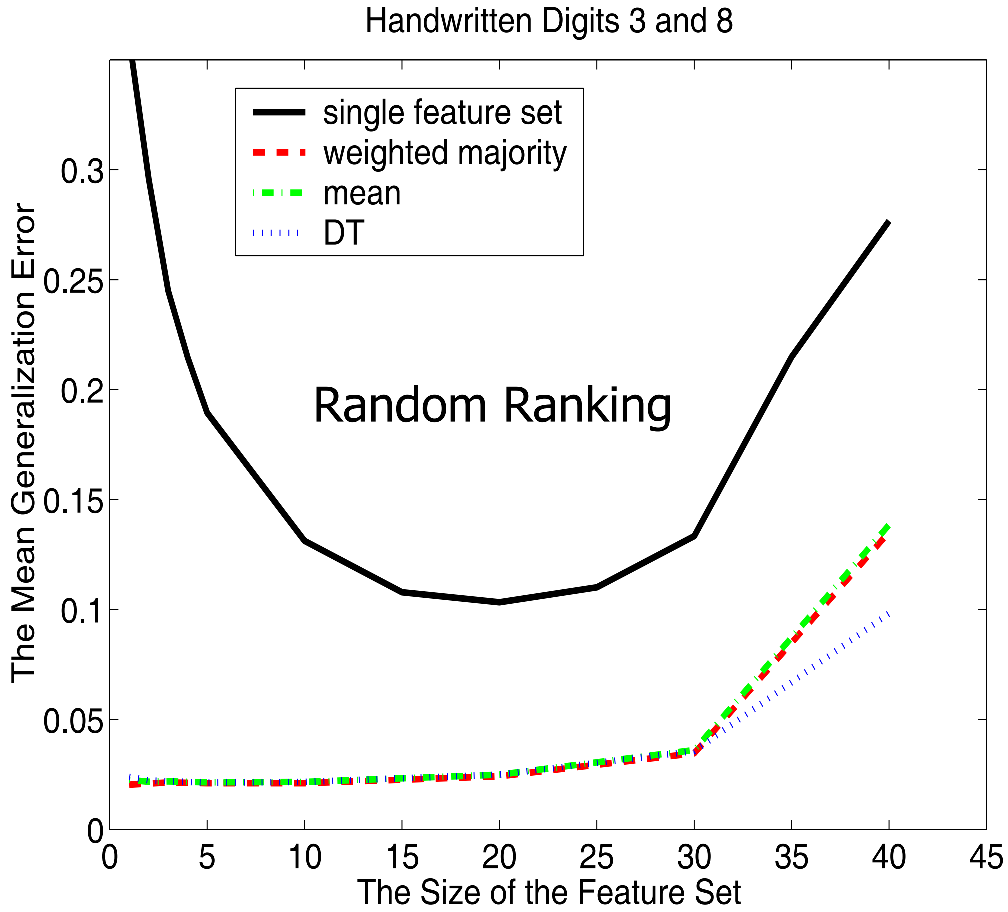
Experiment 1: NIST Digits



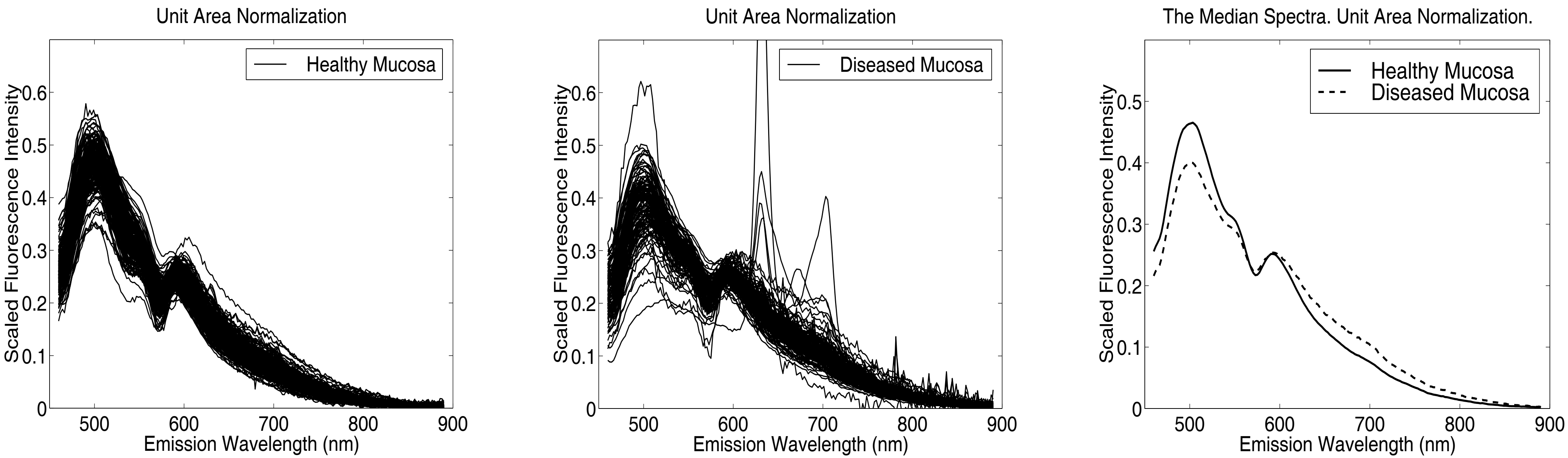
2 x 200 NIST Digits, '3' and '8', 16 x 16 pixels, 256 features.

2 x 100 for training, 2 x 100 for testing





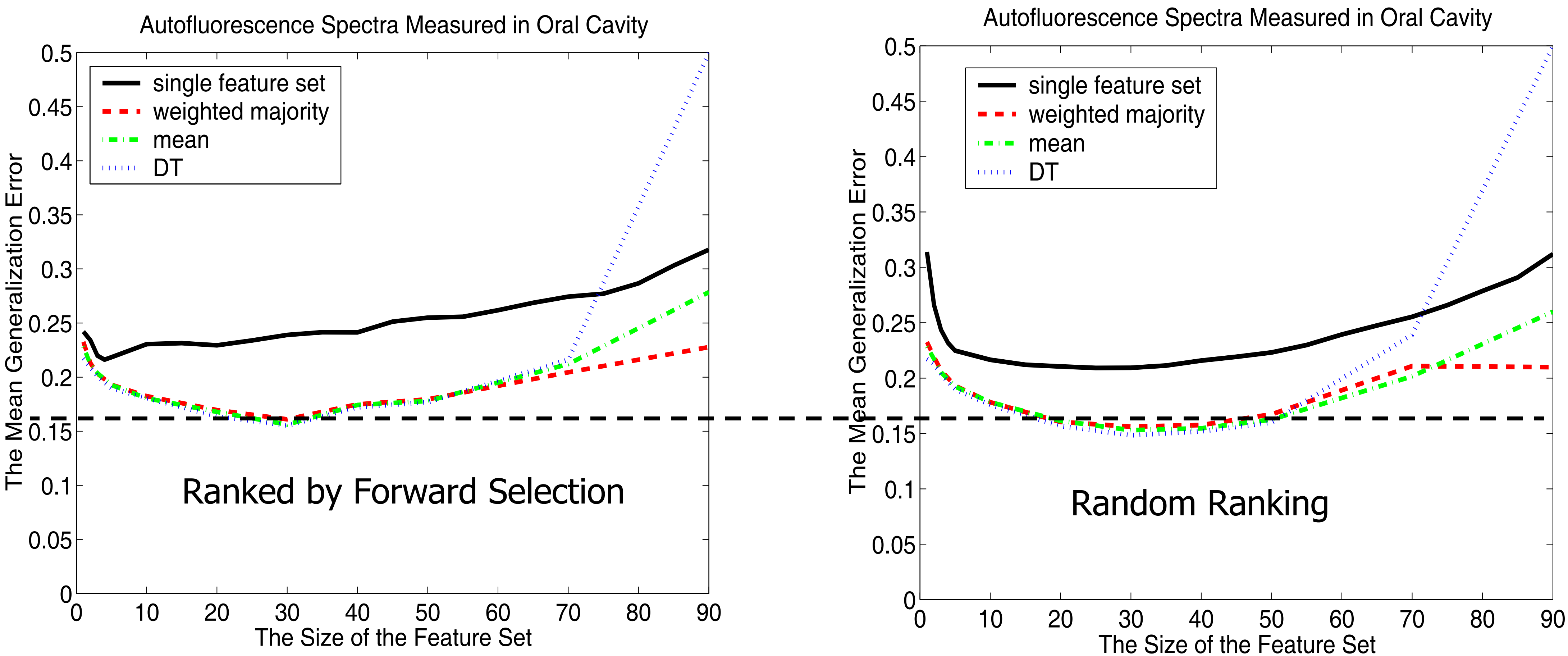
Experiment 2: Spectra



Normalized autofluorescence spectra for healthy and diseased mucosa in oral cavity

Experiment 2: Results

training set, test set, feature size

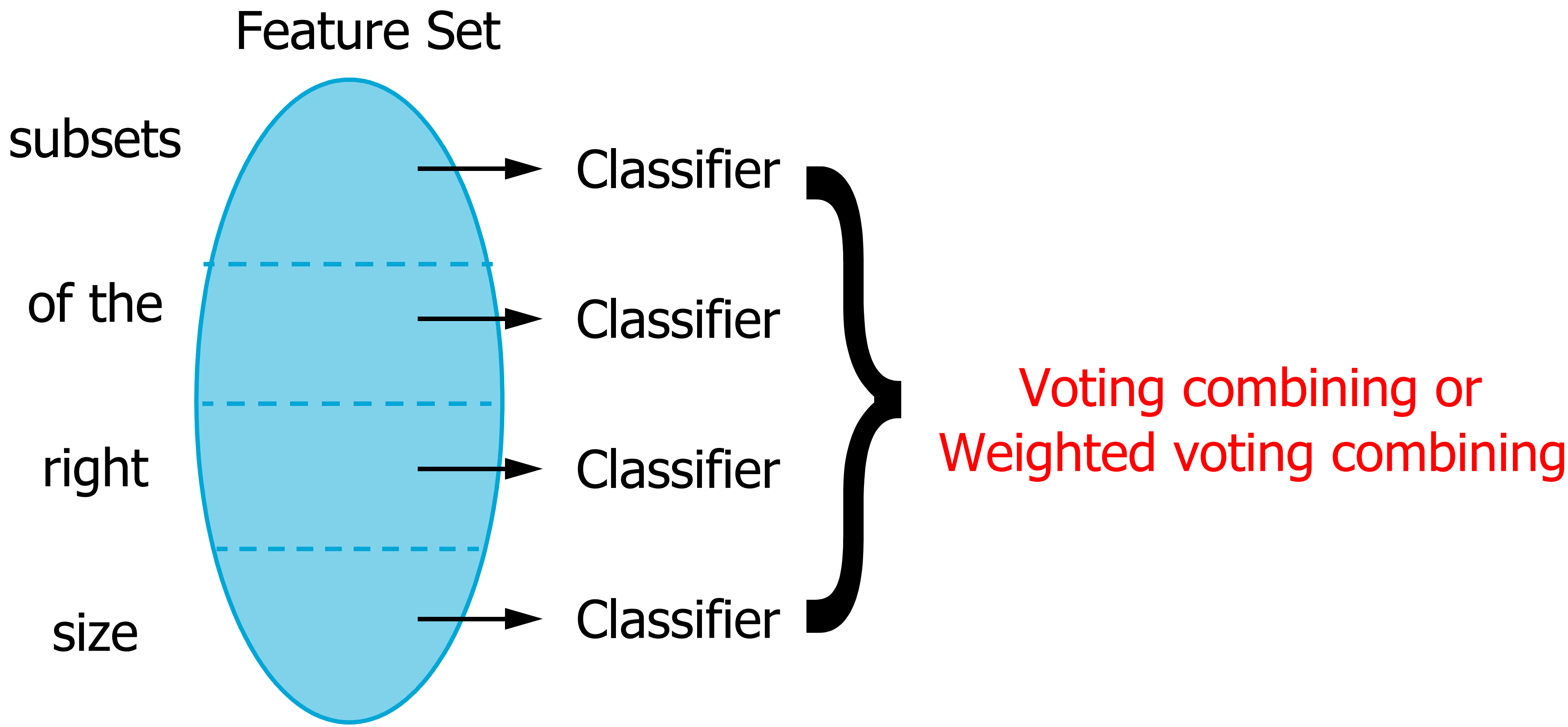


TUDelft

Preliminary Conclusions

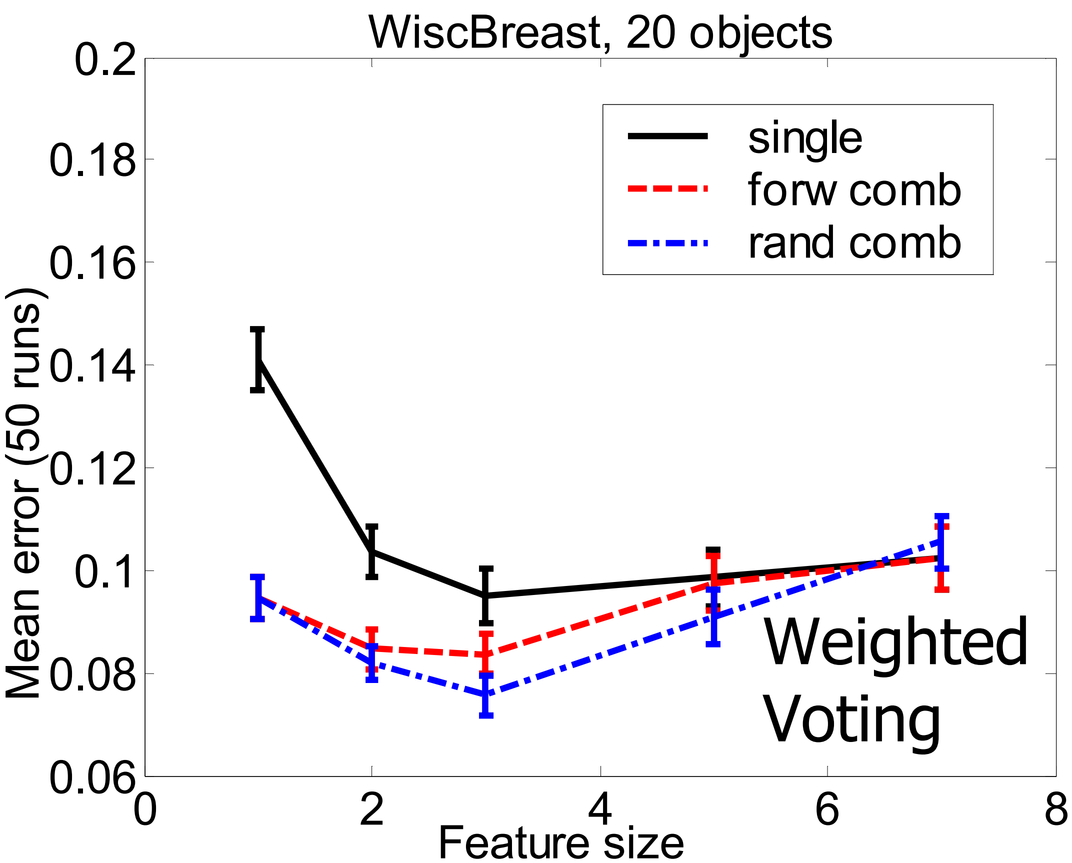
- Subspace combining approach may improve feature selection.
- (Weighted) majority voting seems a good combiner.

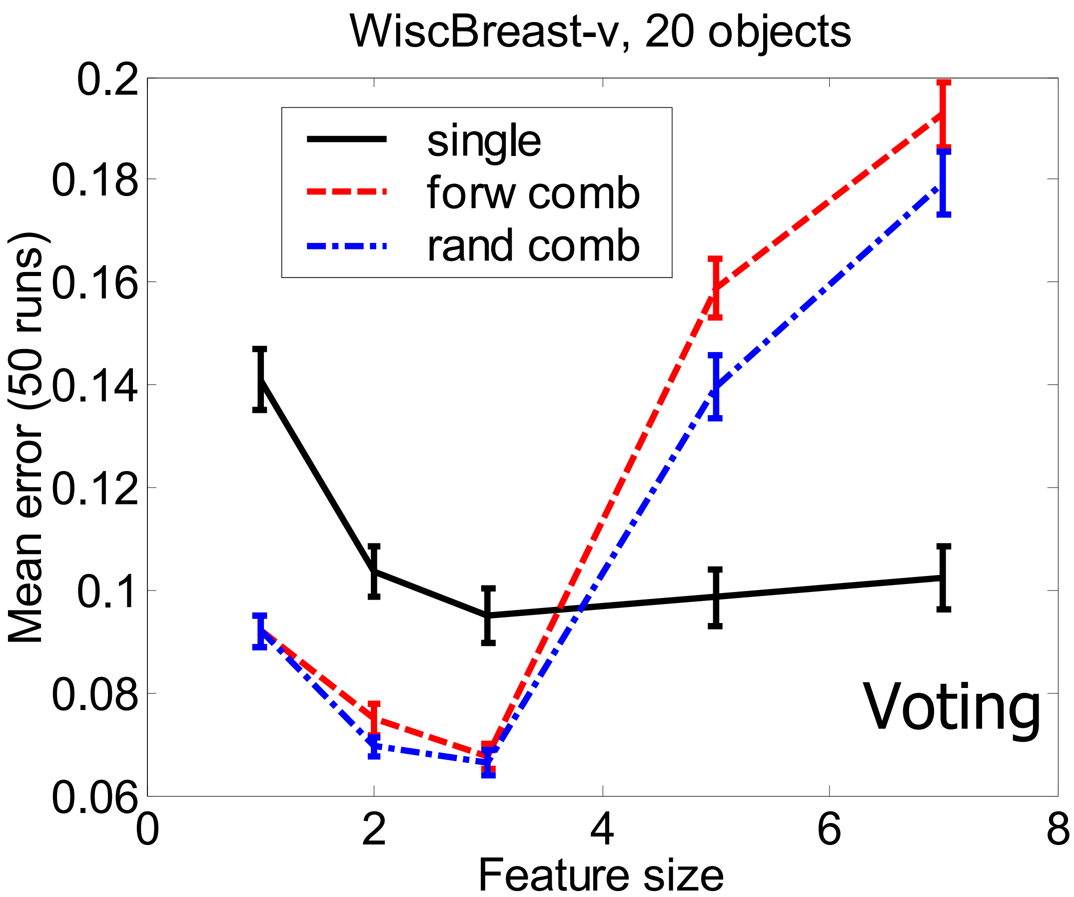
Experimental Verification



Ranking according to forward feature selection
or random ranking ?

Wisconsin Breast Cancer Dataset

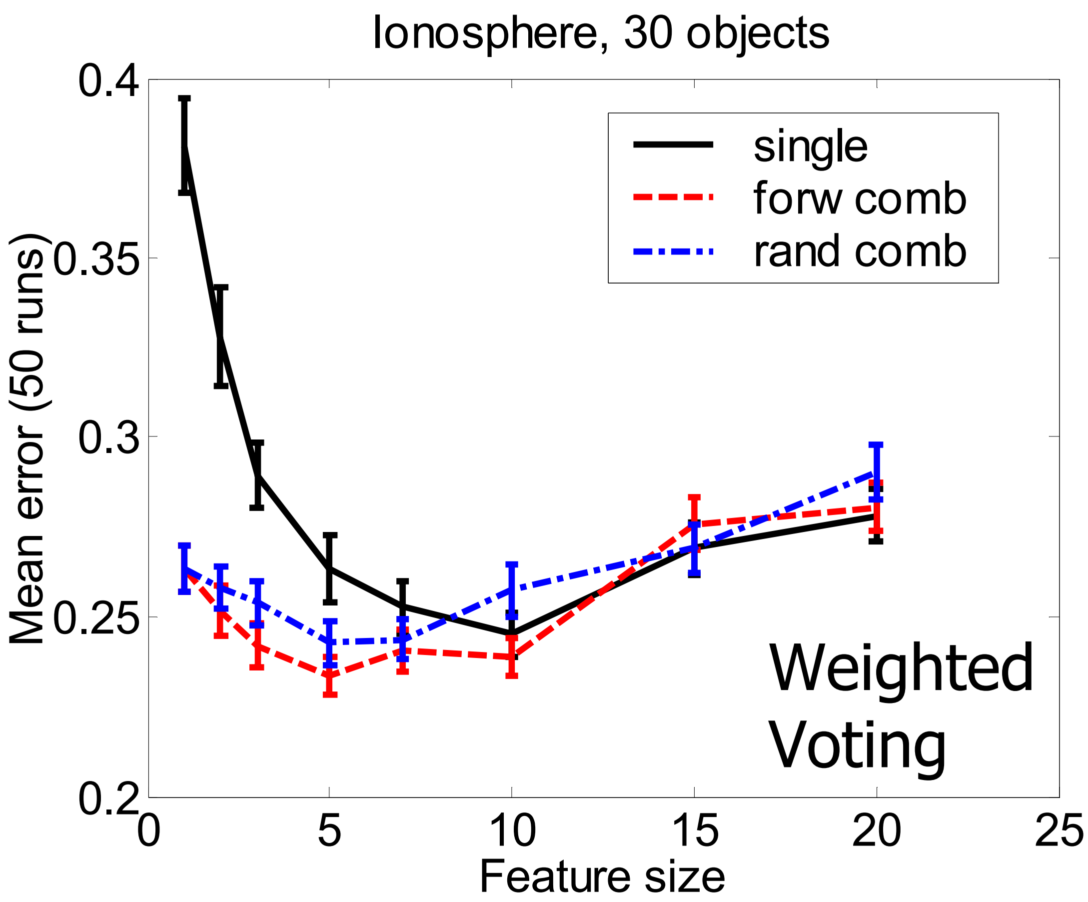


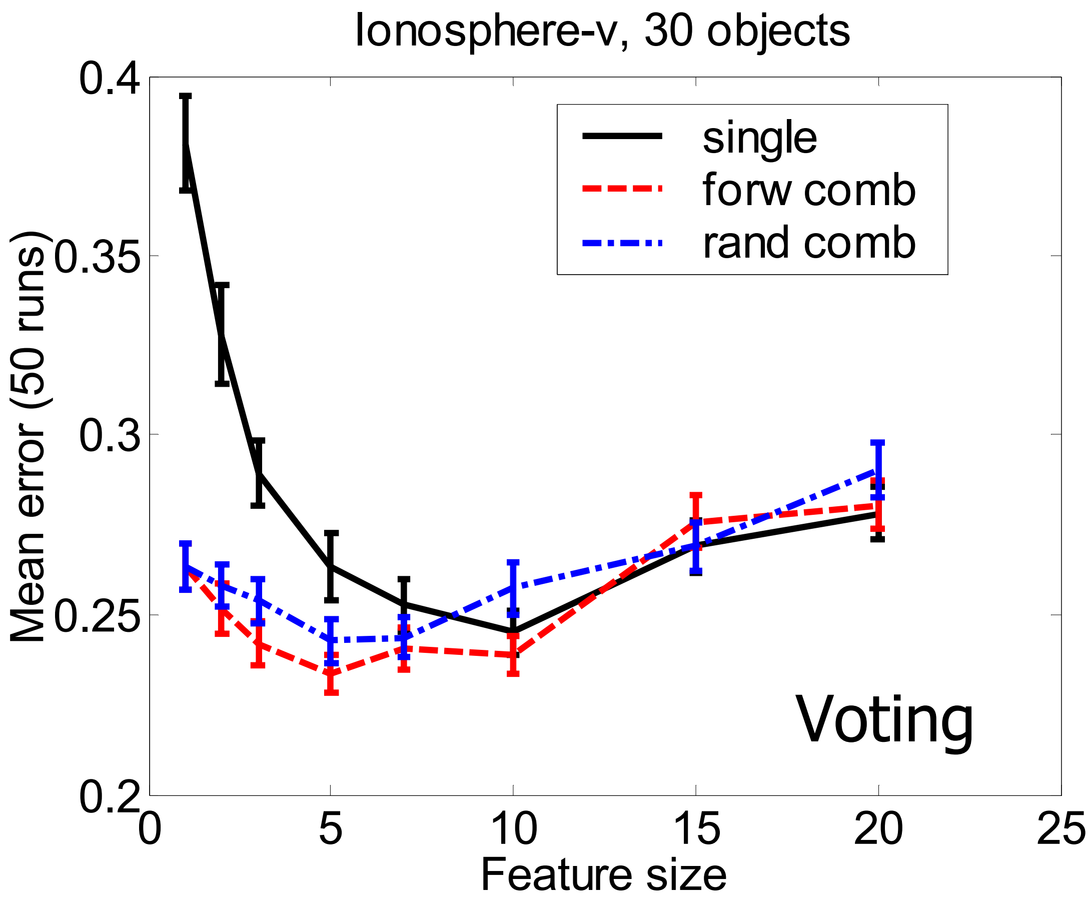


Objects: 683 (444 + 239)

Features: 13

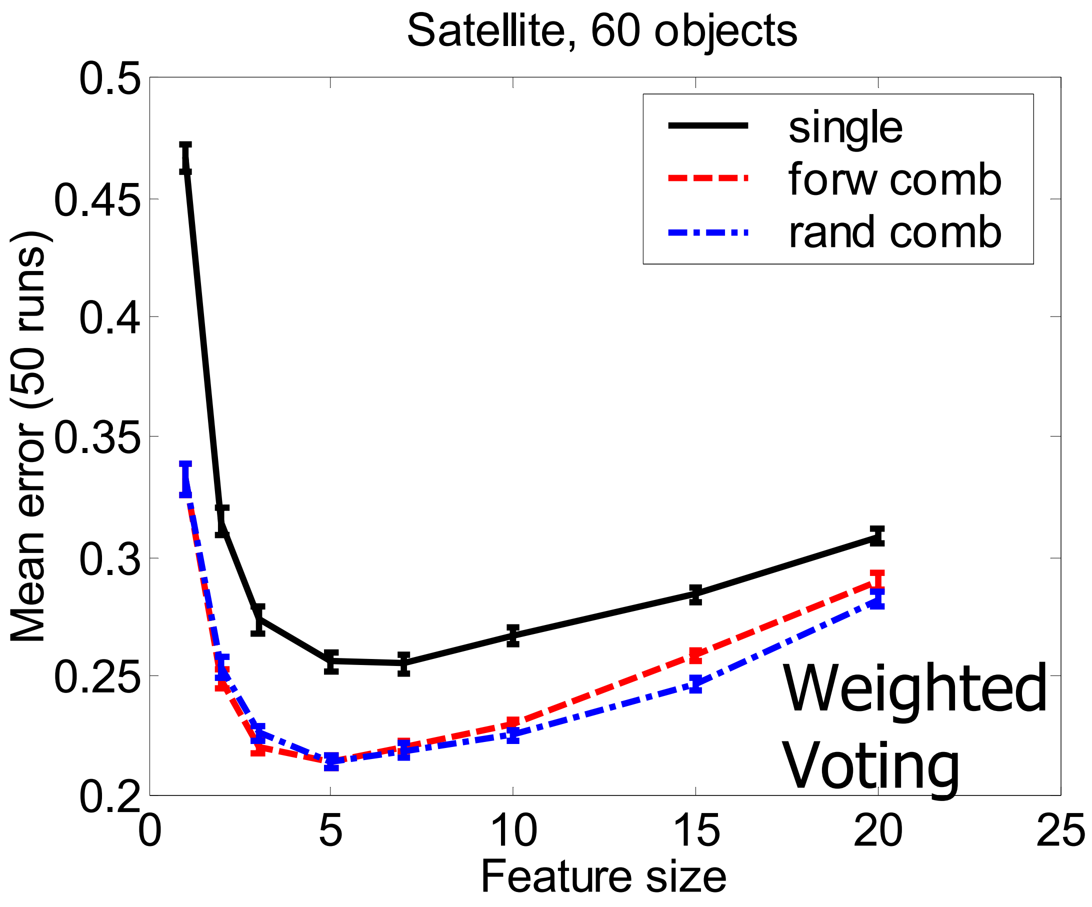
Ionosphere Dataset

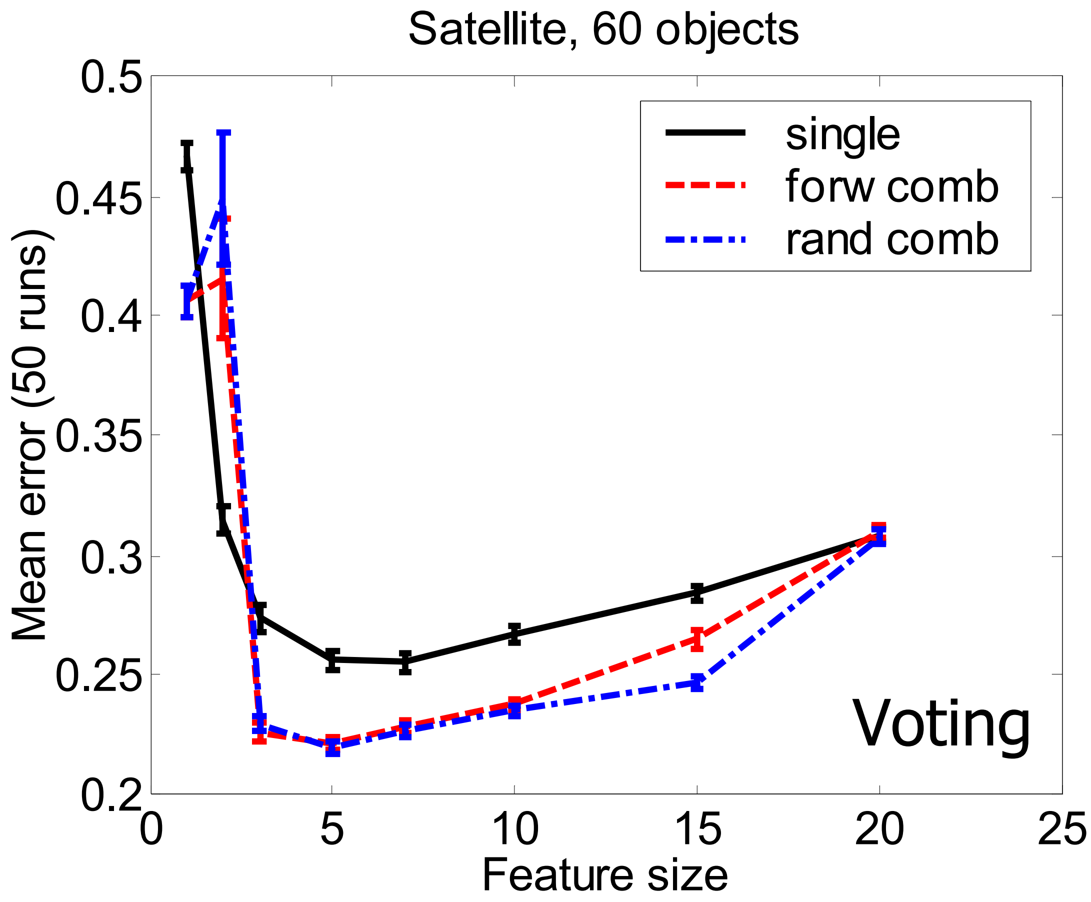




\# Objects: 351 (225+126)

\# Features: 34

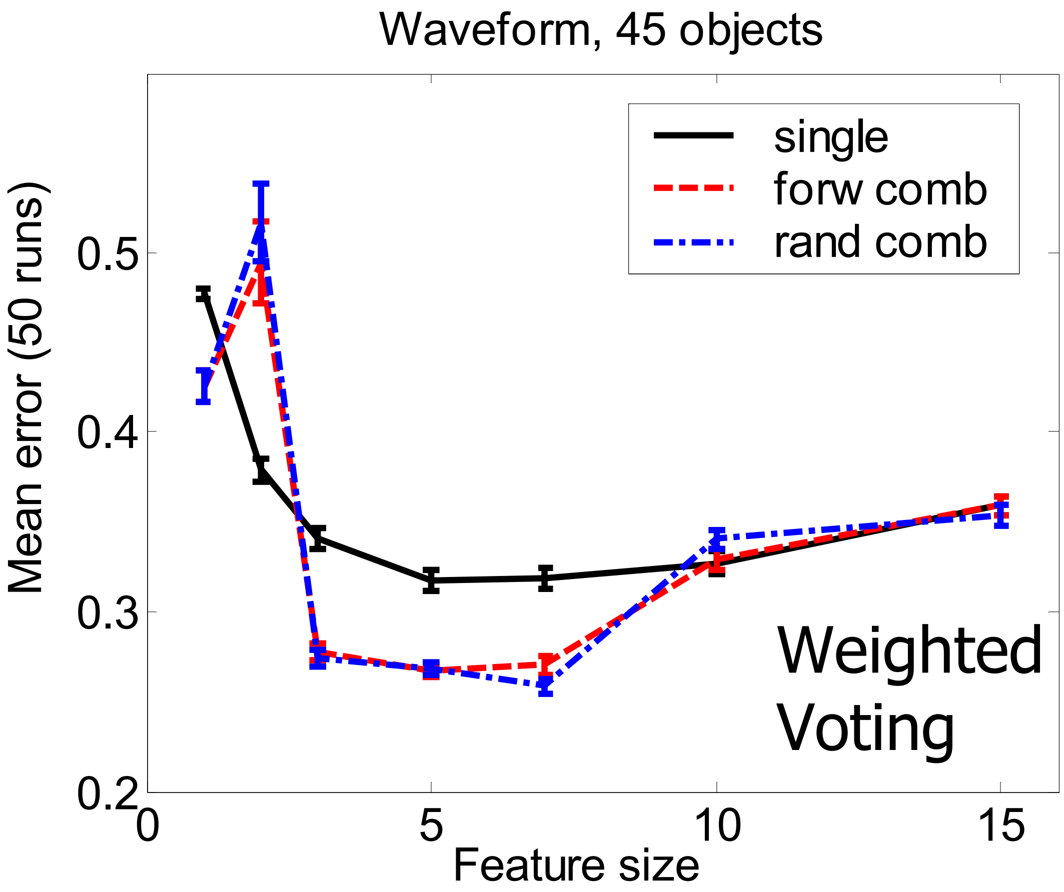


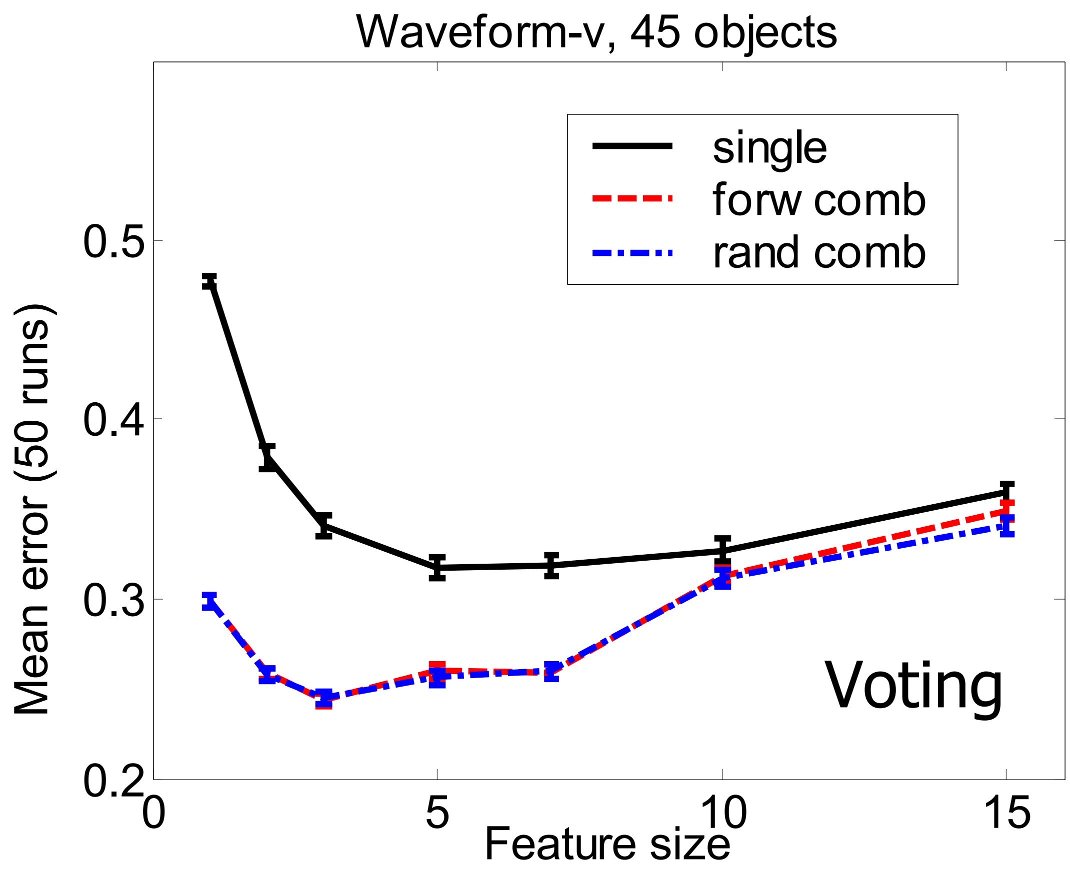


Objects: 6435 (1533+703+1358+626+707+1508)

Features: 36

Waveform



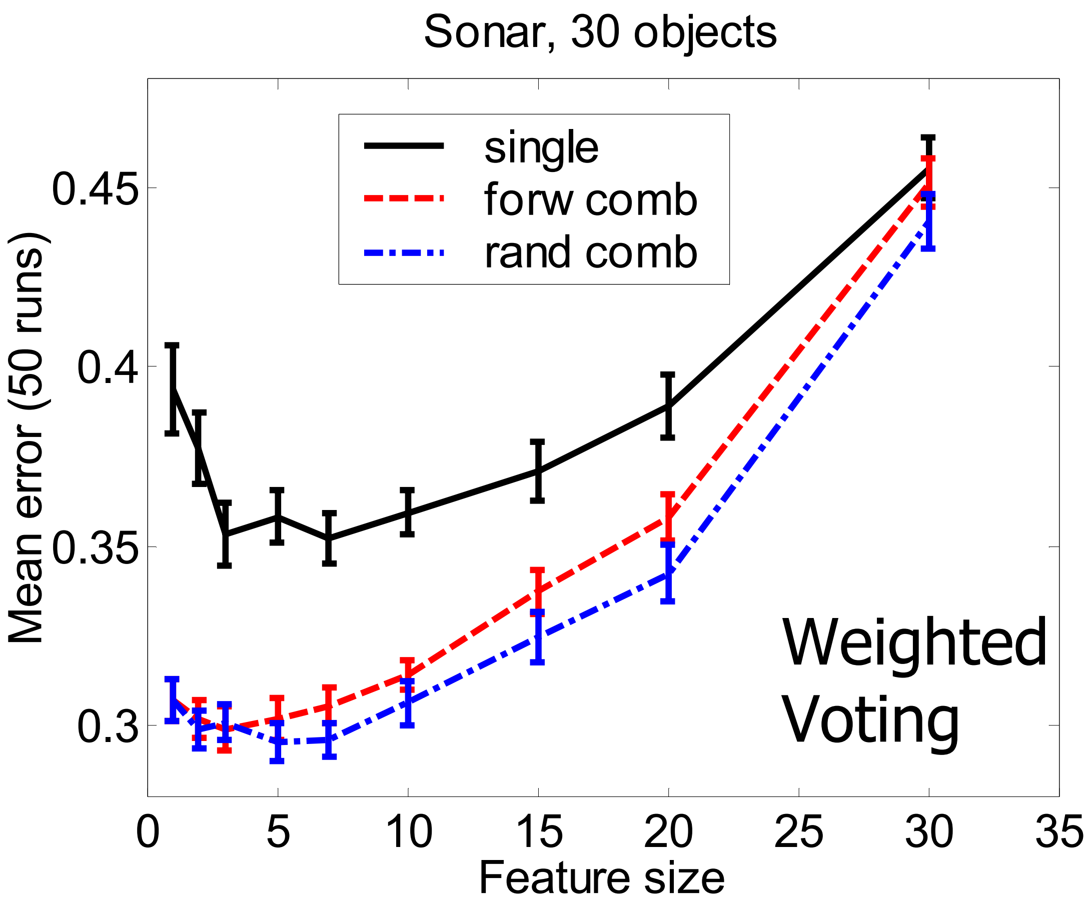


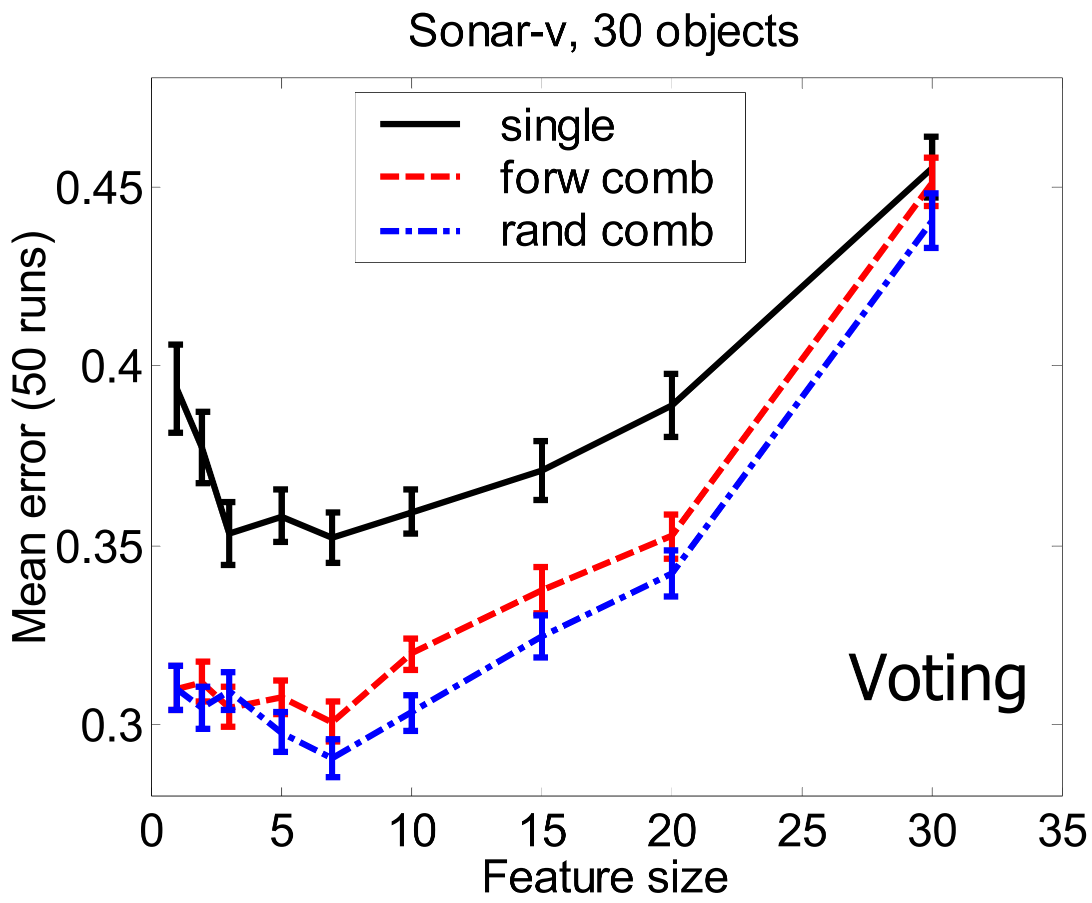
Objects: 5000 (1657+1647+1696

Features: 21

TU Delft

Sonar

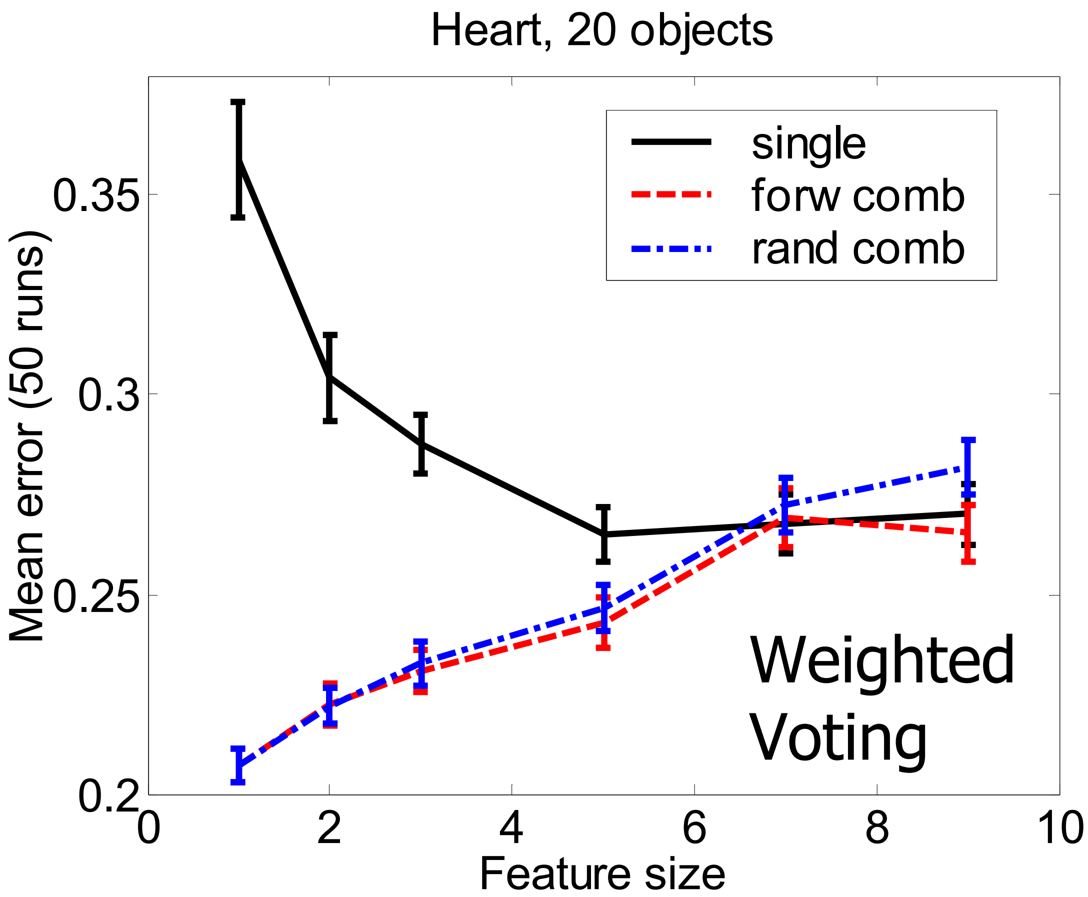


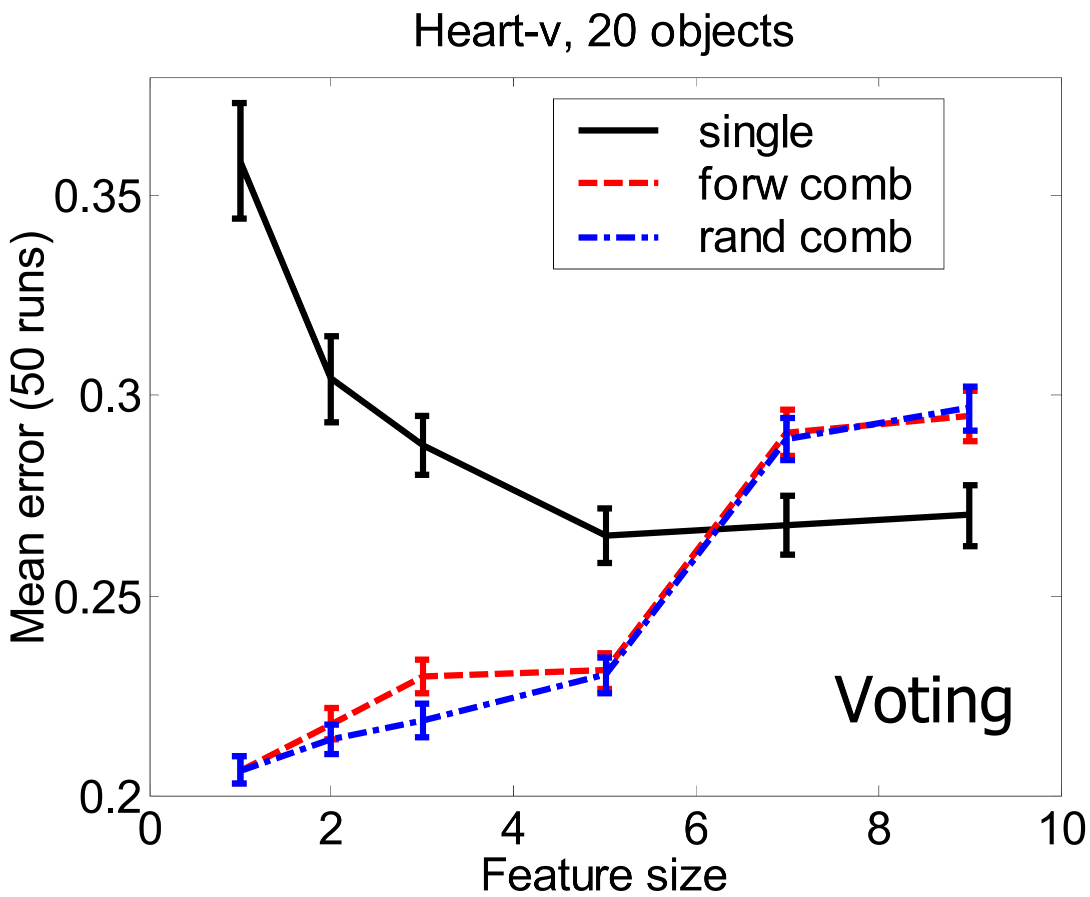


\# Objects: 208 (97+11)

\# Features: 60

Heart Dataset



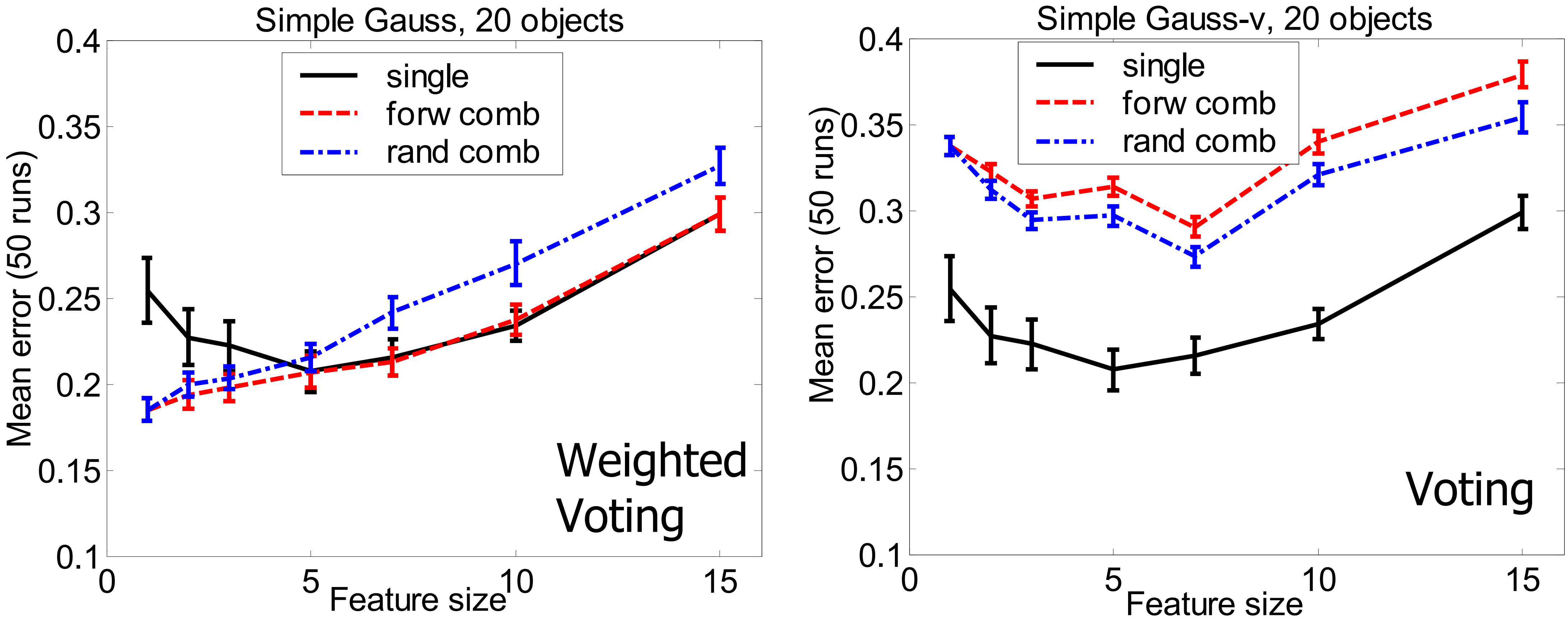


Objects: 297 (160+137)

Features: 13

TU Delft

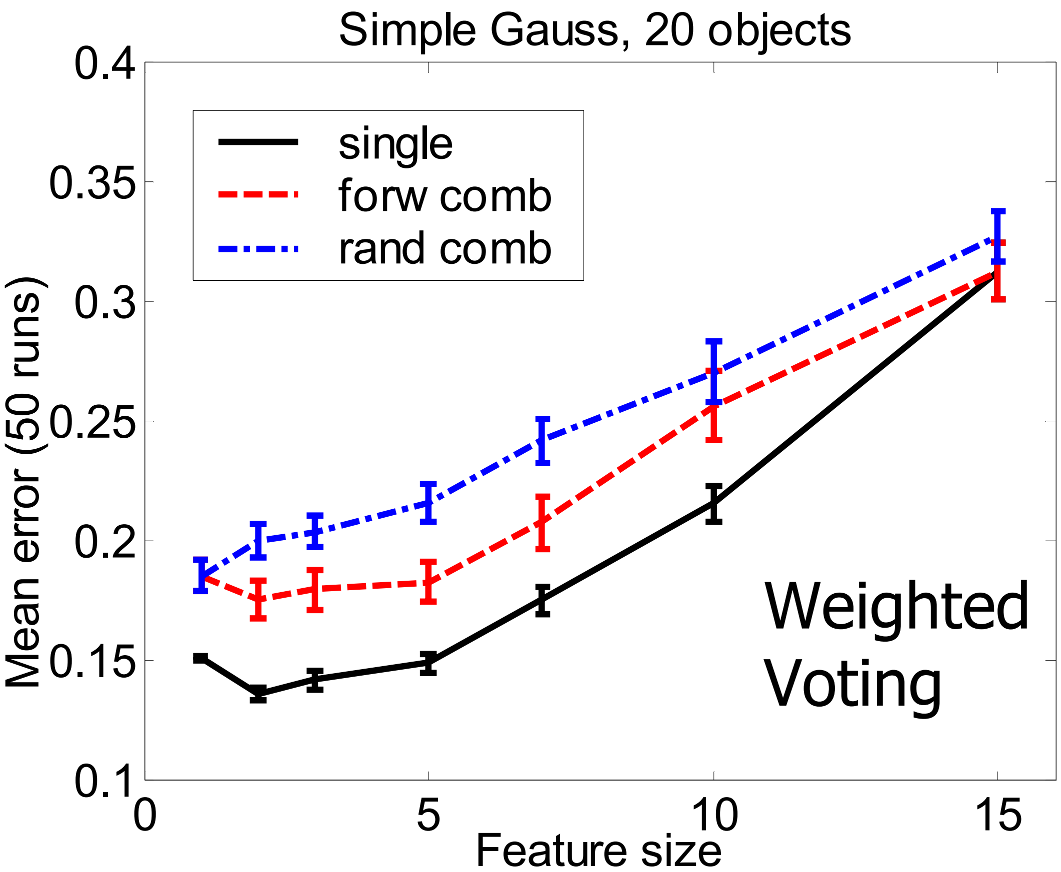
Simple Gauss, No Feature Selection

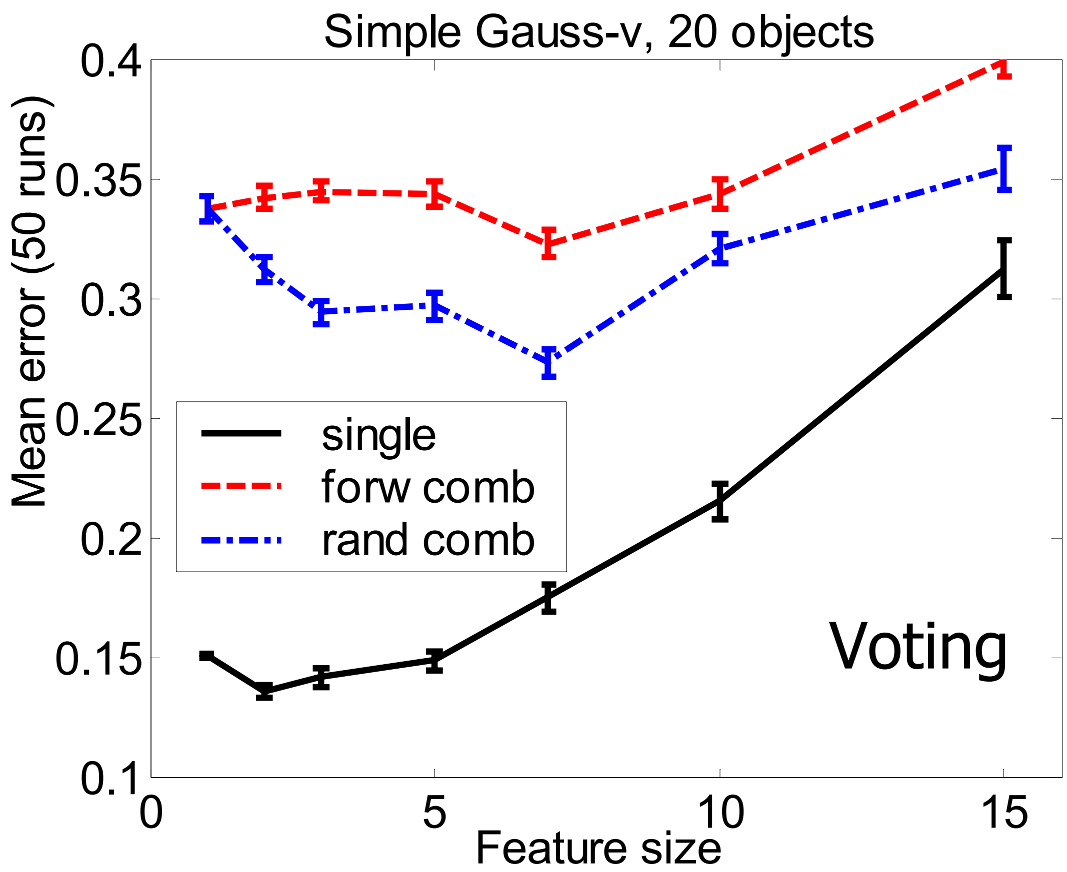


\# Objects: 500 (250+250)

\# Features: 20, d = 2 / n, n = 1, 20

Simple Gauss, Optimal Feature Order





Objects: 500 (250+250)

Features: 20, d = 2 / n, n = 1, 20

TU Delft

Conclusions

- Feature selection may be improved by a combined subspace approach based on all features.
- This approach does not deteriorate the result.
- Random feature ordering might do equally well.

Conclusions

- Feature selection may be improved by a combined subspace approach based on all features.
- This approach does not deteriorate the result.
- Random feature ordering might do equally well.
- The procedure may not work for an initially well ranked set of features