Improving Cross-Validation Classifier Selection Accuracy through Meta-learning

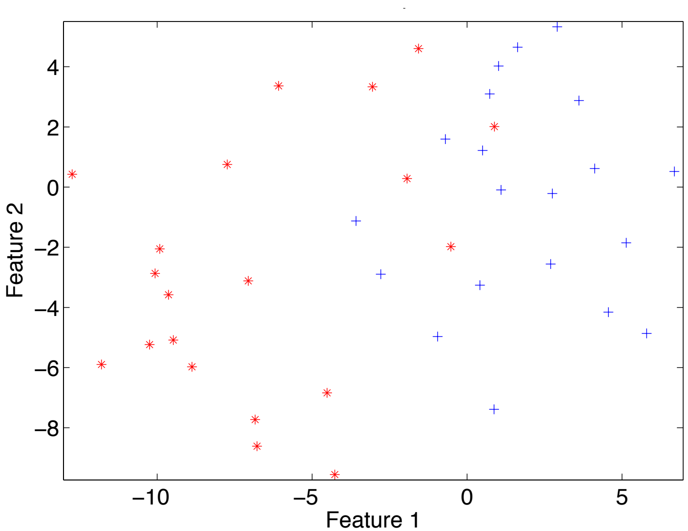
Jesse H. Krijthe, Tin Kam Ho, Marco Loog

COMMIT/
LUMC Leiden University Medical Center
Alcatel·Lucent
TU Delft Delft University of Technology

Classifier Selection Problem

Classification problem

D



Classifiers

SVM
QDA
LDA
GBM
Random Forest
Fisher
Nearest Mean
Nearest Neighbor
C4.5
ID3

Which classifier gives the lowest error rate e when evaluated on a large test set?

A Practical Solution

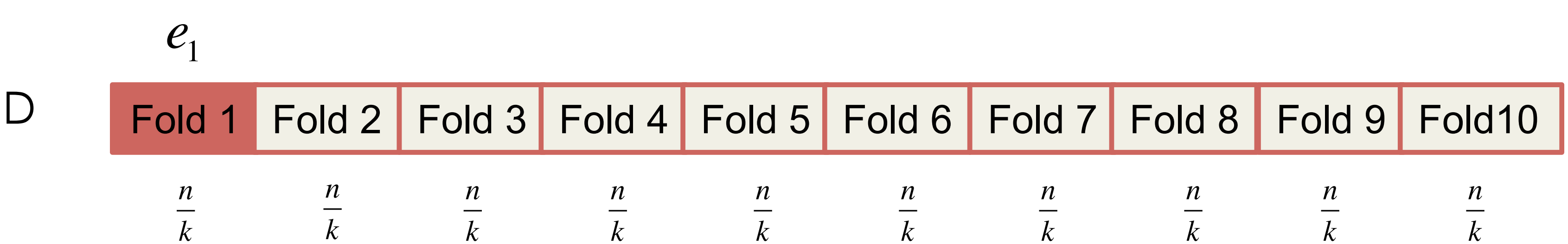
- In practice: have no large test set to determine e
- Alternative: estimate e through a cross-validation procedure, $\hat{e}_{cv}$
- Procedure is *practically* unbiased and intuitive
- Use the estimates of each classifier to select the best one
- Used for:
  - Classifier selection
  - Parameter tuning
  - Feature selection
  - Performance estimation

Goal

Is it possible to use meta-learning techniques to improve the *accuracy* (rather than the computational efficiency) of classifier selection using cross-validation?

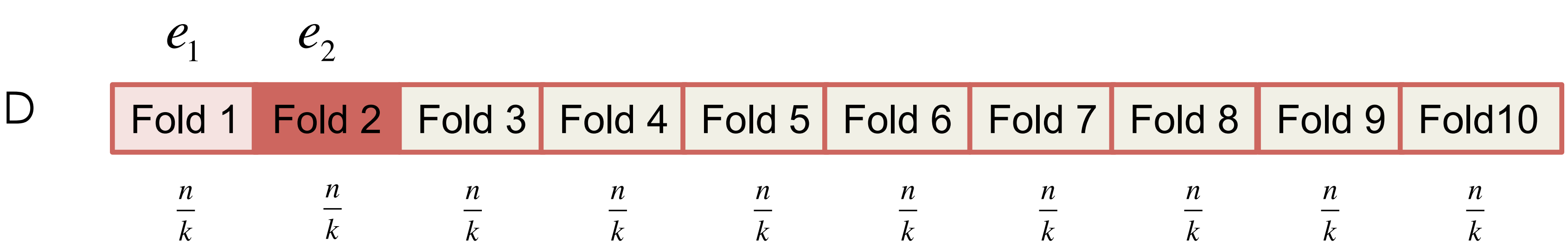
Cross-validation revisited (1/2)

- C={c_1,..c_m} a set of classifiers, D a dataset
- Calculate the *k*-fold cross-validation error
  1. Randomly assign the *n* objects in the dataset to *k* parts (folds)
  2. Use fold 2 to *k* to train a classifier
  3. Use fold 1 to test its accuracy
  4. Cycle through, using each fold as the test set once
  5. Average the accuracies over all the folds



Cross-validation revisited (1/2)

- C={c_1,..c_m} a set of classifiers, D a dataset
- Calculate the *k*-fold cross-validation error
  1. Randomly assign the *n* objects in the dataset to *k* parts (folds)
  2. Use fold 2 to *k* to train a classifier
  3. Use fold 1 to test its accuracy
  4. Cycle through, using each fold as the test set once
  5. Average the accuracies over all the folds



Cross-validation revisited (1/2)

- $C=\{c_1,..c_m\}$ a set of classifiers, D a dataset
- Calculate the *k*-fold cross-validation error
  1. Randomly assign the *n* objects in the dataset to *k* parts (folds)
  2. Use fold 2 to *k* to train a classifier
  3. Use fold 1 to test its accuracy
  4. Cycle through, using each fold as the test set once
  5. Average the accuracies over all the folds

$$\hat{e}_{cv} = \sum_{i=1}^{k} \frac{e_i}{k}$$

	e_1	e_2	e_3							
D	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold10
	$\frac{n}{k}$	$\frac{n}{k}$	$\frac{n}{k}$	$\frac{n}{k}$	$\frac{n}{k}$	$\frac{n}{k}$	$\frac{n}{k}$	$\frac{n}{k}$	$\frac{n}{k}$	$\frac{n}{k}$

Cross-validation revisited (2/2)

- Select the classifier with lowest $\hat{e}_{cv}$
- Bias decreases as *k* increases
  - Unbiased as estimator for $n - \frac{n}{k}$
  - Small bias for reasonable *k*, large *n*

- For a particular dataset, interested in the difference e and $\hat{e}_{cv}$
- Variance
  - High as *k* goes to *n*
  - High as *k* goes to 2
  - Lowest usually around 5-10
  - Higher than for bootstrap and resubstition

Why would cross-validation fail?

- As Braga-Neto et. al. 2004 and others note, if ***n*** is small, variance of the cross-validation error estimate becomes large
- Cross-validation error estimates become unreliable for a given dataset
- Specifically: classifier selection based on these estimates may suffer

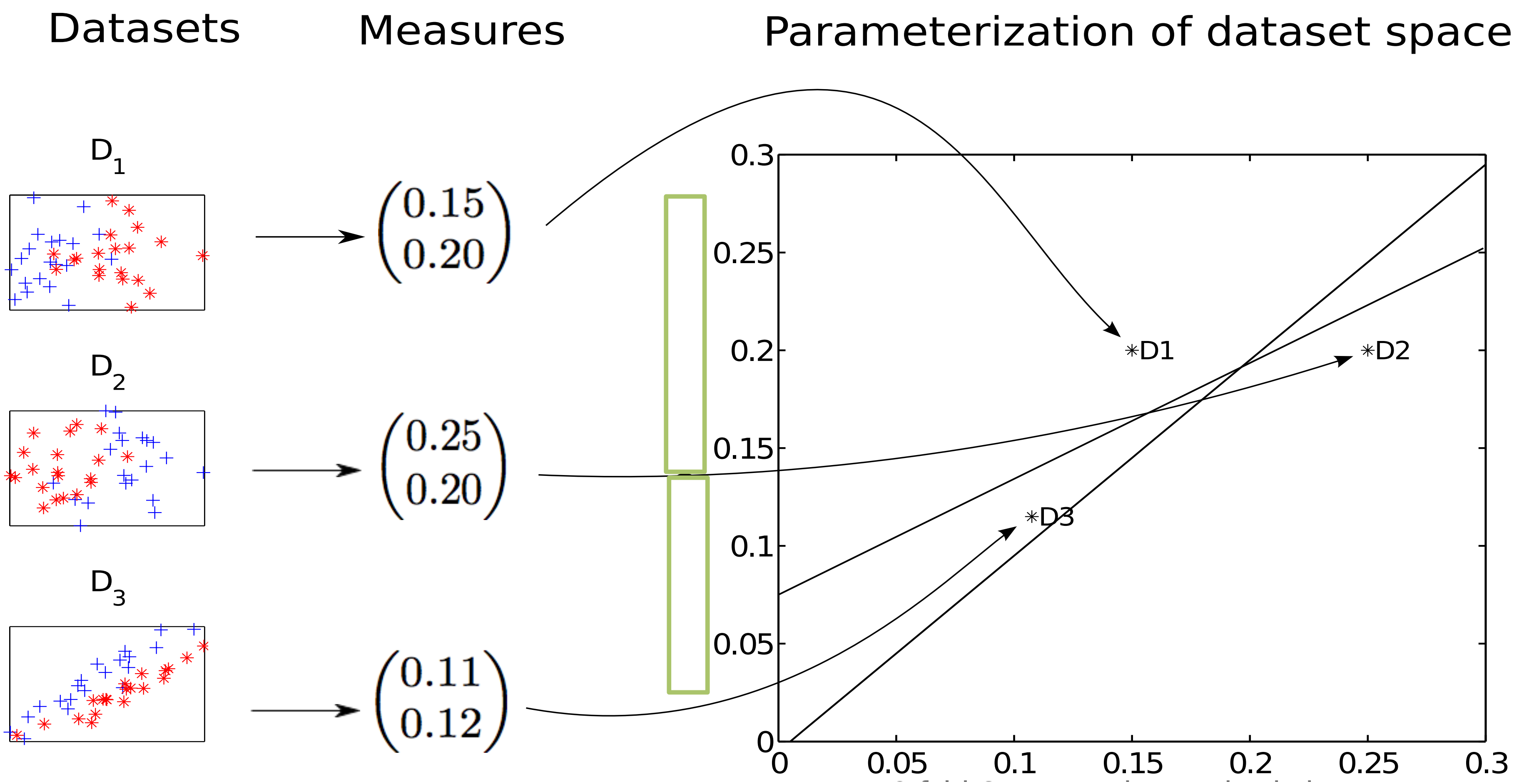
Meta-learning (1/2)

- Learning which classifier to select based on characteristics of the dataset

- Classifier selection as just another classification problem
  - Classes: the most accurate classifier
  - Features: statistics on the dataset (meta-features)

- Meta-features are preferably
  - Computationally efficient
  - Predictive
  - Interpretable

Meta-learning (2/2)

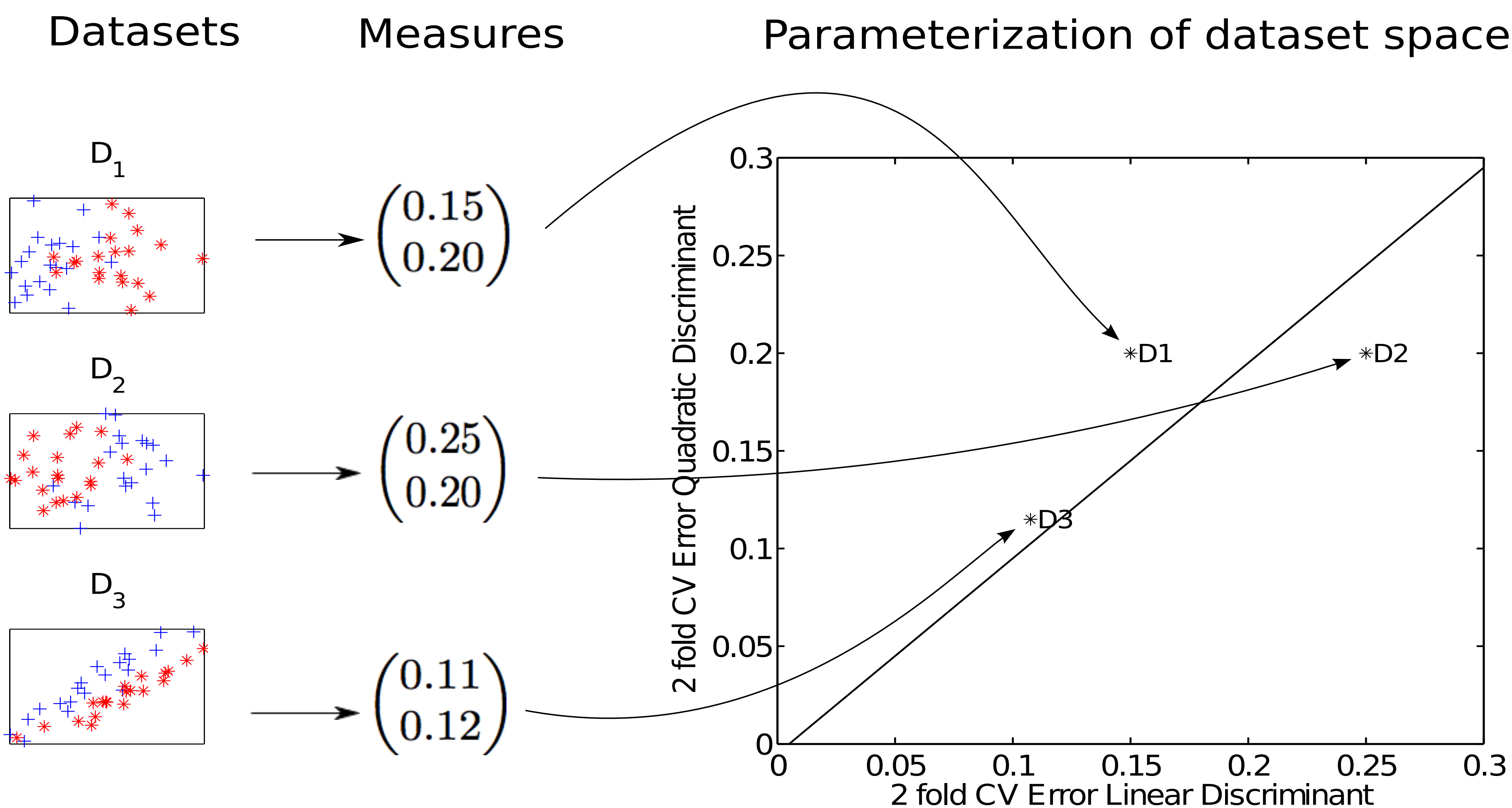


Cross-validation selection as meta-learning

- Cross-validation errors are measures on the dataset as well
- Idea: Treat them as meta-features
- Meta-classifier in this case:
  - Select the classifier with the lowest cross-validation error
  - Static diagonal rule

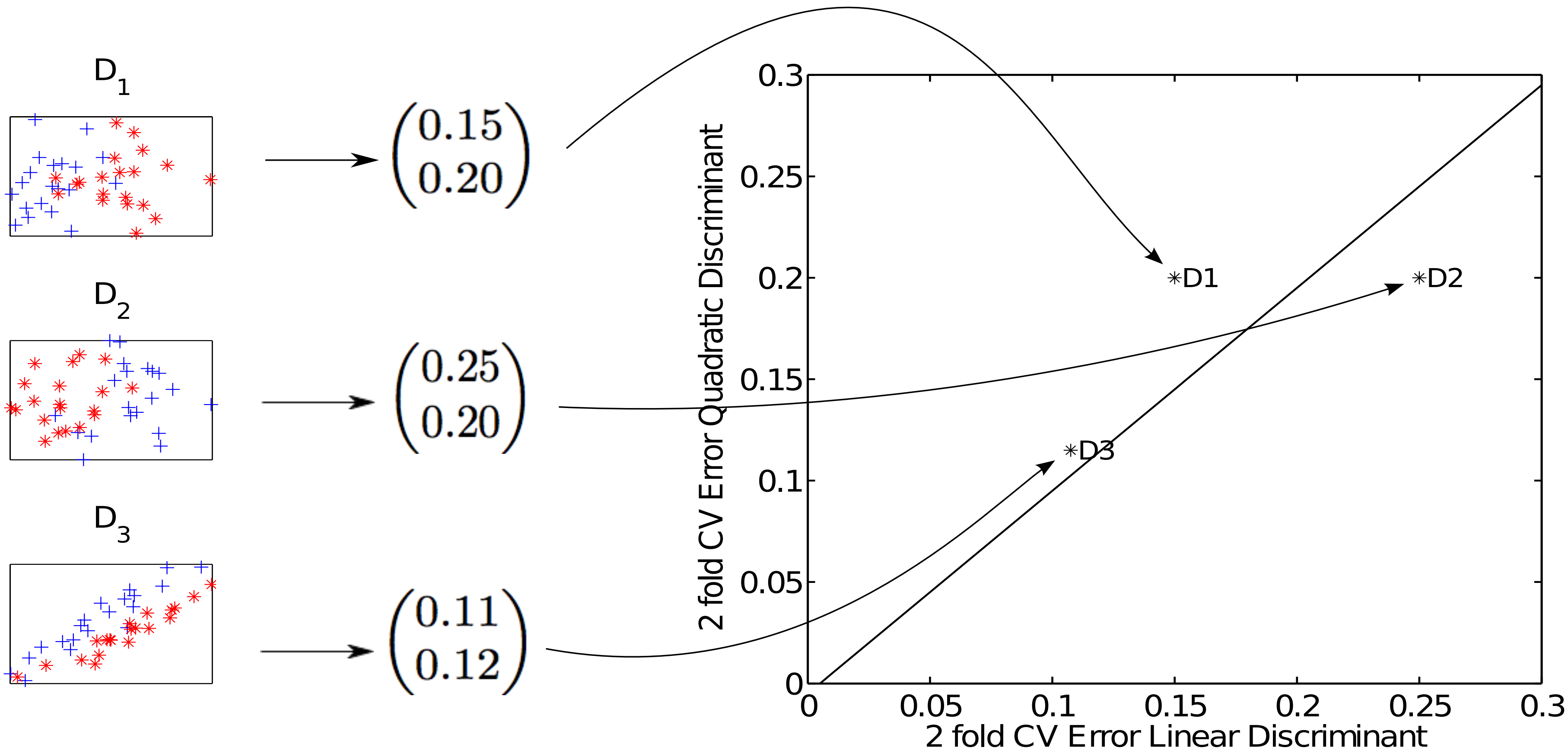
Meta-classes	Best classifier (m)
Meta-features	Cross-validation error (m)
Meta-classifier	Static ‘diagonal’ rule

Cross-validation Meta-problem



Cross-validation Meta-problem

Datasets Measures Parameterization of dataset space



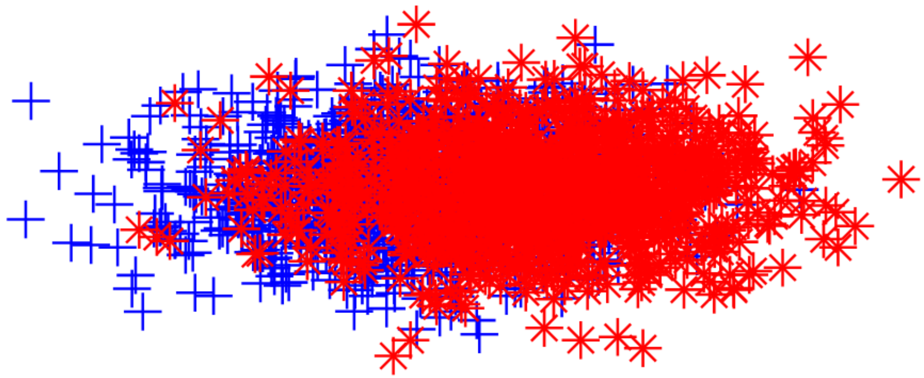
Is this simple, static rule justified?

A Meta-learning Universe (1/3)

- Choice between two simple classifiers:
  - Nearest Mean
  - 1-Nearest Neighbor
- Two simple problem types
  - Each suited to one of the classifiers
  - Small training samples (20-100)
  - Generate enough data to estimate the real error (~20000)
  - Problem types have equal priors
- Slightly contrived
  - Visualization
  - Illustrate Concept

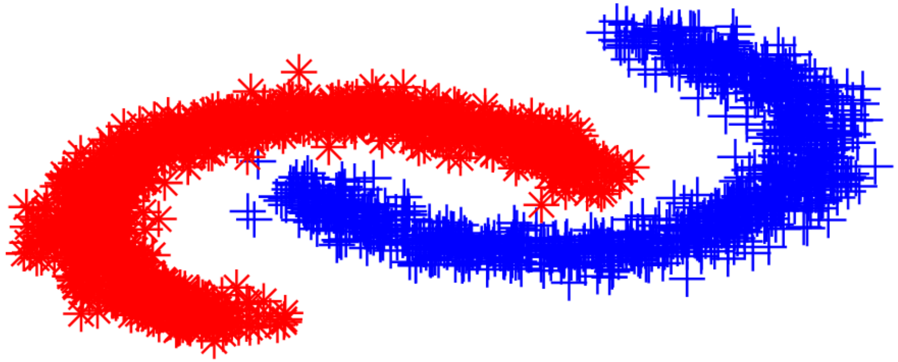
A Meta-learning Universe (2/3)

Gaussian problem

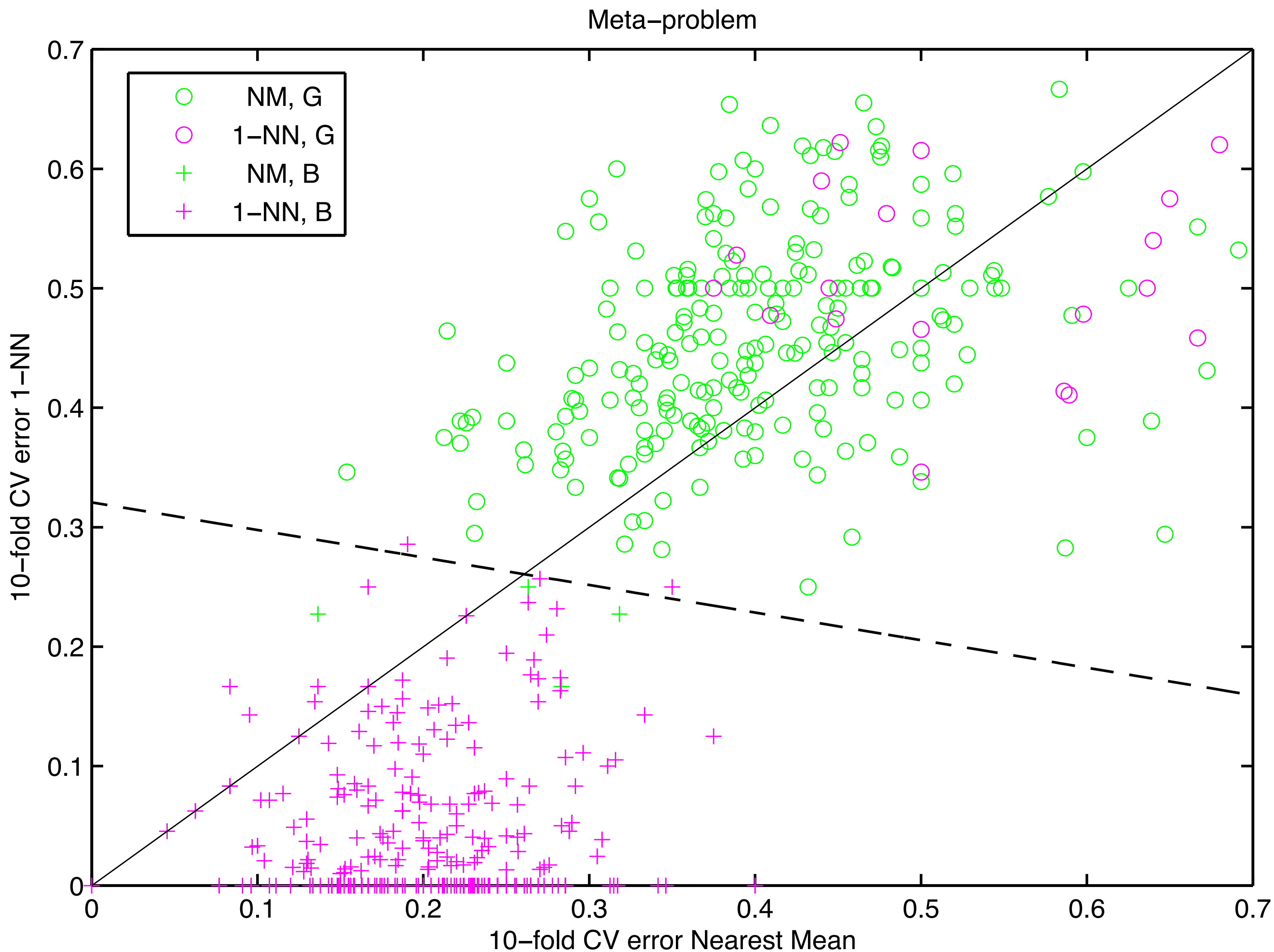


- Randomly vary the distance
- Generate 500 problems
- $G=\{G_1,G_2,...,G_{500}\}$
- High Bayes error

Banana Set



- Randomly vary the width (variance)
- Generate 500 problems
- $B=\{B_1,B_2,...,B_{500}\}$
- Low Bayes error

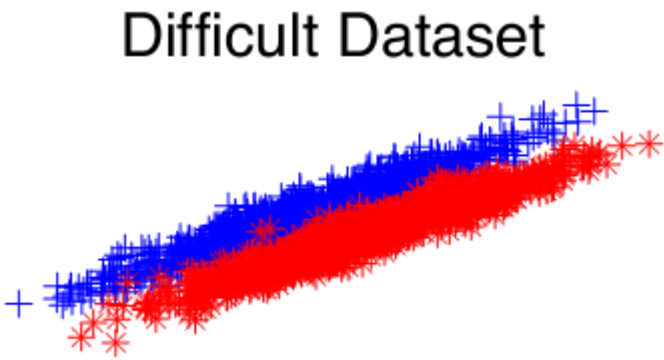


Error: 0.16 -> 0.06 (learning makes a difference)

Additional meta-features (1/2)

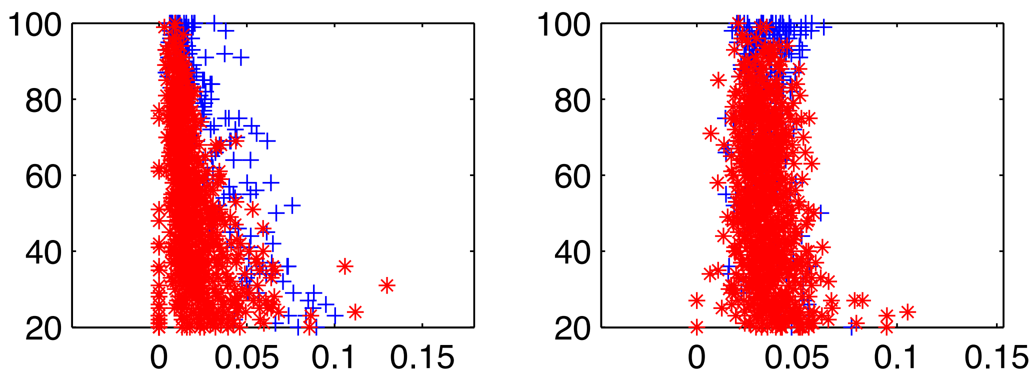
Can characteristics of the data improve classifier selection after we know the cross validation errors?

- Classifiers: Nearest mean and Least Squares
- Elongated boundary problem (100 dimensions)
- Randomness
  - Class priors
  - Number of objects (20-100)
- Extra features
  - Number of objects *n*
  - Variance of the cross-validation errors

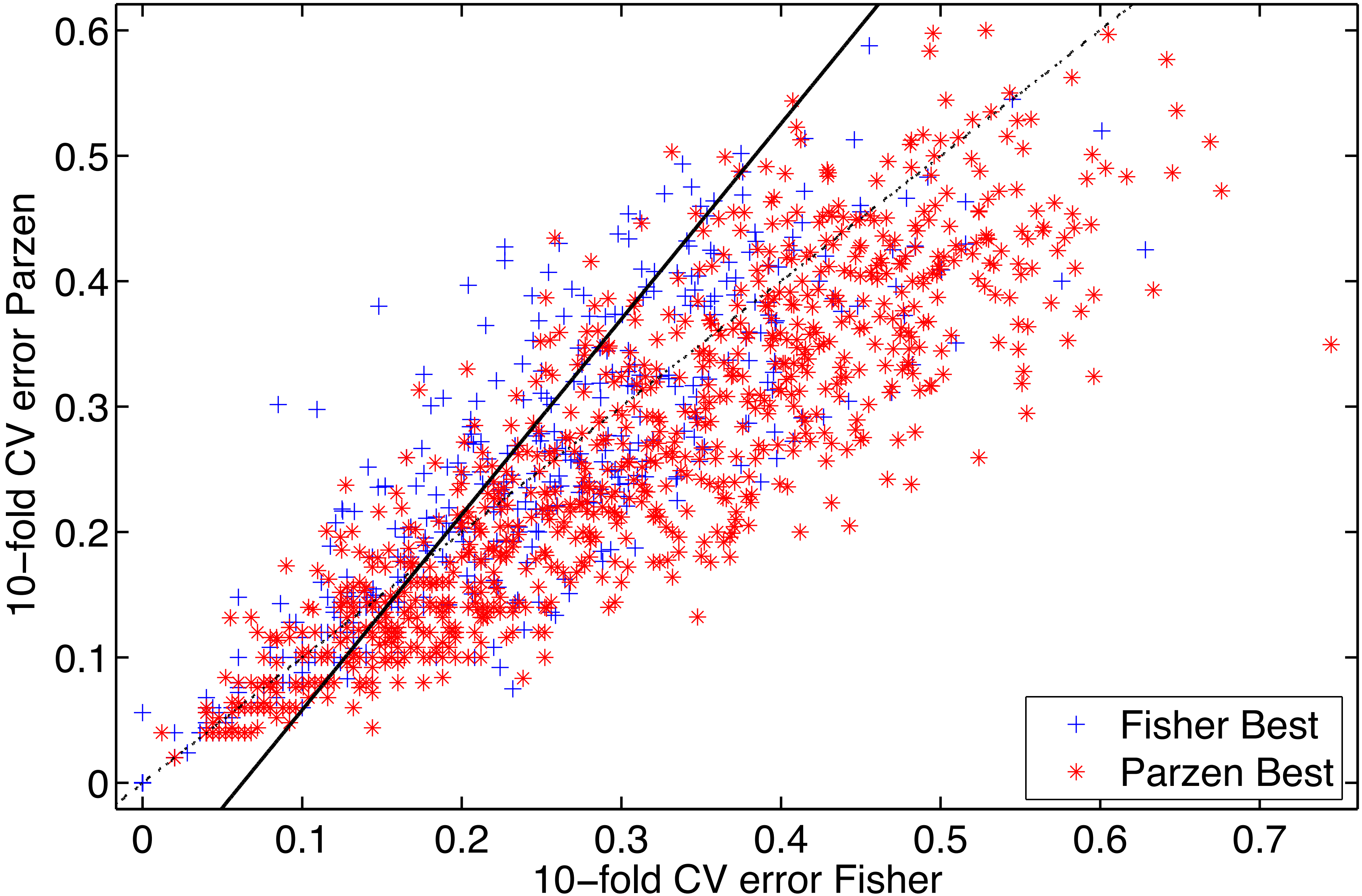


Additional meta-features (2/2)

Classifier	CV errors	+n	+Variance	+n & Variance
CV-selection	0.237			
k-NN	0.238	0.151	0.221	0.127
LDA	0.241	0.159	0.239	0.110



Pseudo real-world data



Pseudo real-world data

Classifier	Best on
Nearest Mean	236
k-Nearest Neighbor	118
Fisher	243
Quadratic Discriminant	32
Parzen Density	286
Decision Stump (Purity Criterion)	221
Linear Support Vector Machine	164
Radial Basis Support Vector Machine	200

Classifier	CV errors	+Variance
CV-selection	0.695	
k-NN	0.605	0.587
LDA	0.618	0.599

Conclusion

- There are universes were **meta-learning can outperform cross-validation** based classifier selection
- Additional statistics of the data can aid in classifier selection
- Some indication this works on real-world datasets, more experiments are needed
- Evidence to support meta-learning not just as a time-efficient alternative to cross-validation, but potentially more accurate