Women in Machine Learning Workshop
12th of December 2011

Ina Fiterau, Carnegie Mellon University
Artur Dubrawski, Carnegie Mellon University

Explaining Datasets through High-Accuracy Regions

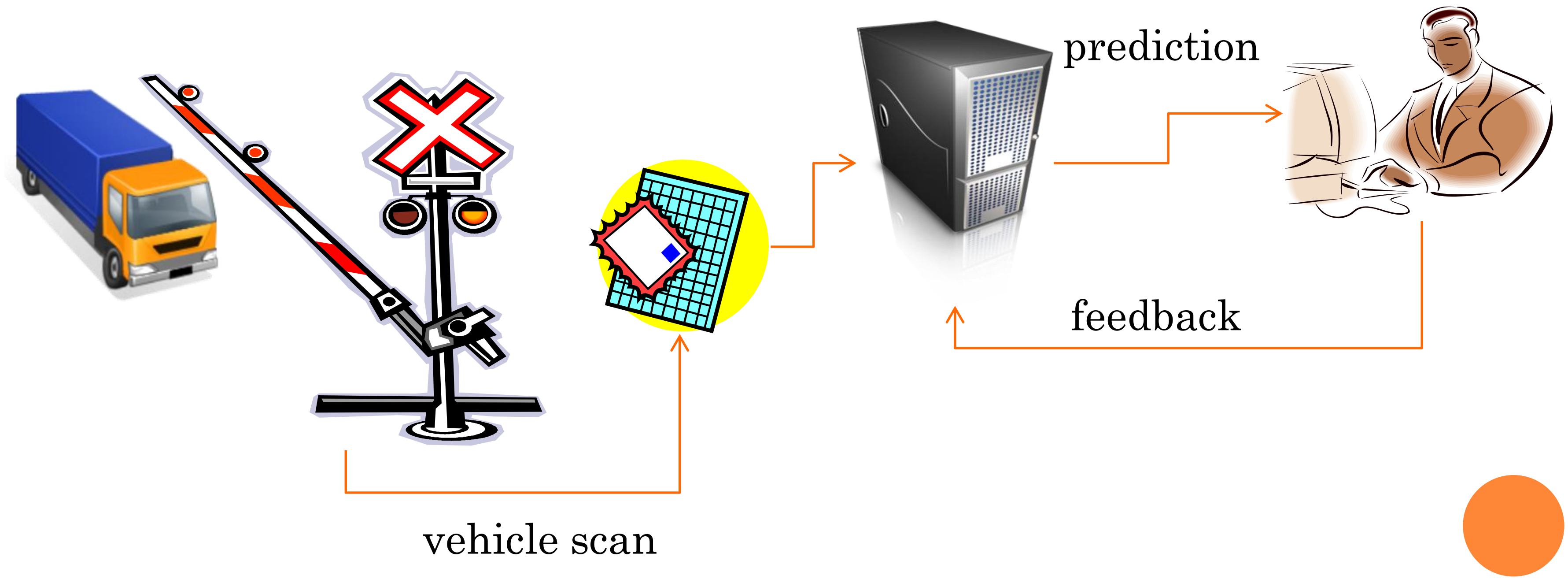
Work under review at the SIAM Data Mining Conference

Outline

- Motivation of need for interpretability
- Explanation-Oriented Partitioning (EOP)
- Evaluation of EOP

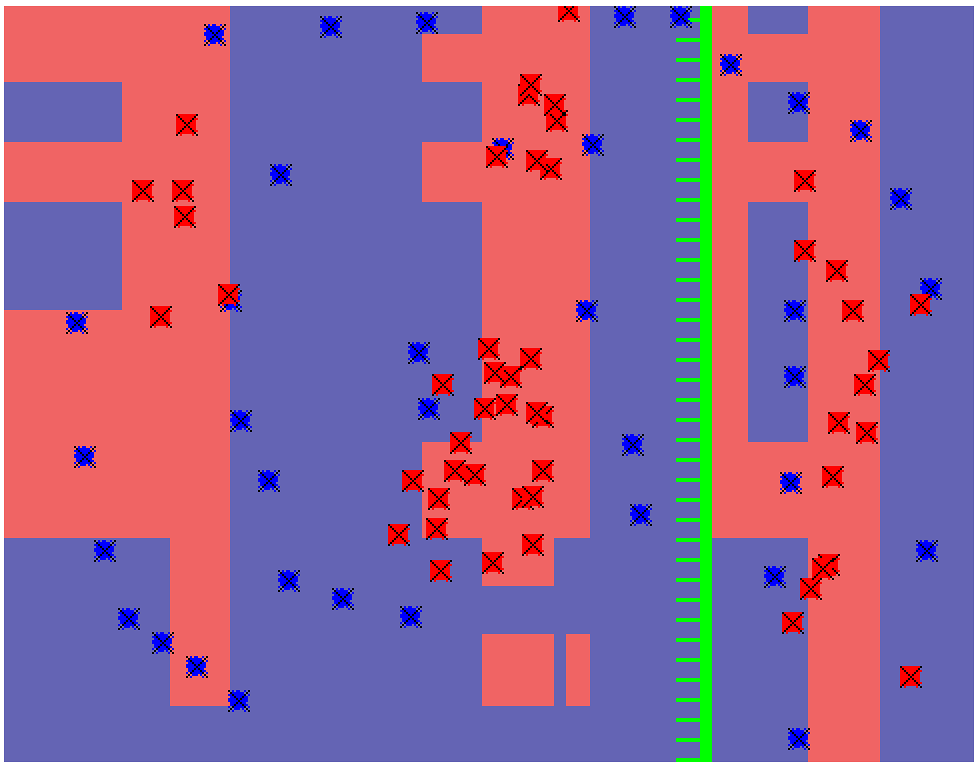
Example Application: Nuclear Threat Detection

- Border control: vehicles are scanned
- Human in the loop interpreting results



Boosted Decision Stumps

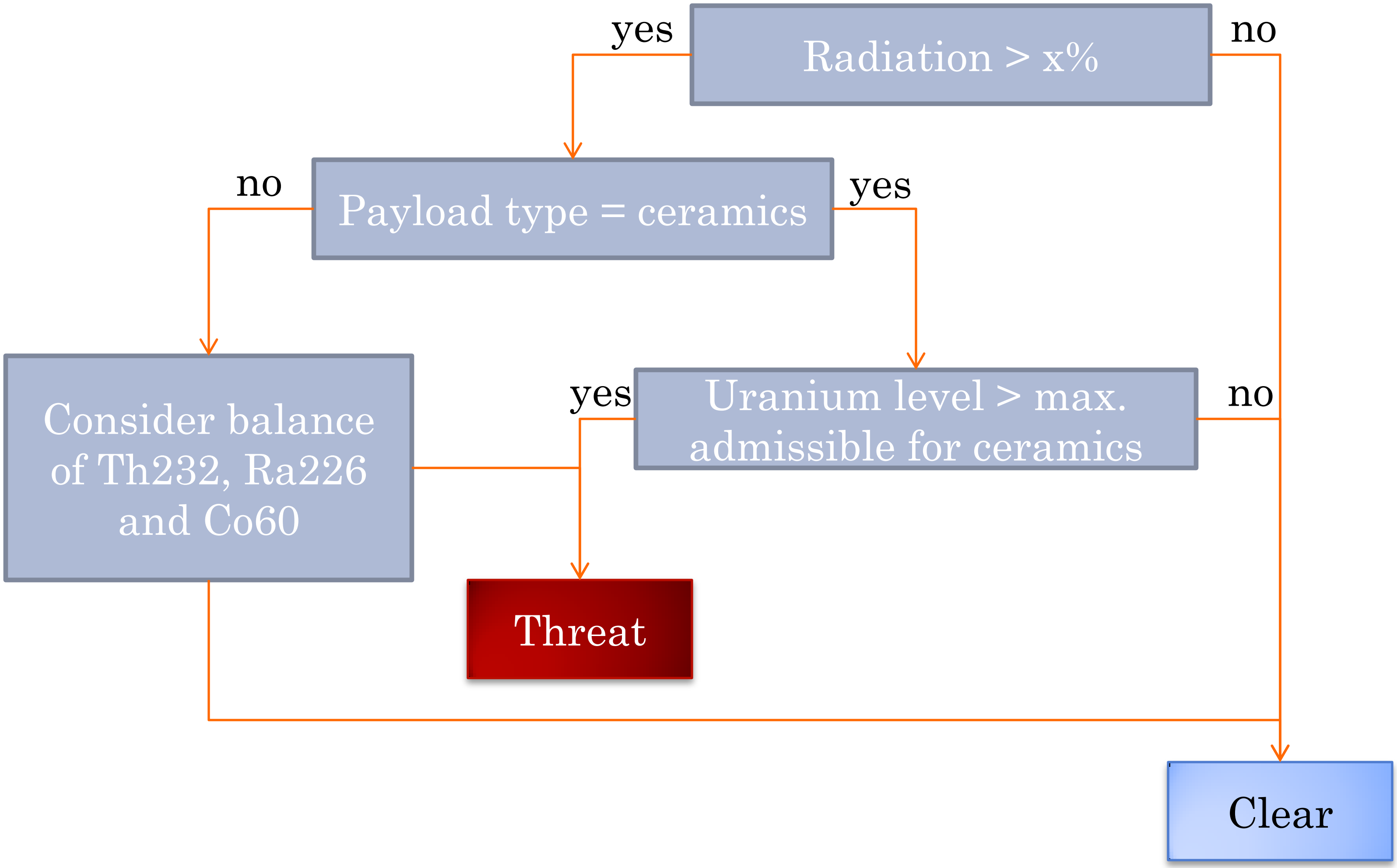
- Accurate, but hard to interpret



How is the prediction derived from the input?

Image obtained with the Adaboost applet.

Decision Tree – More Interpretable



Motivation

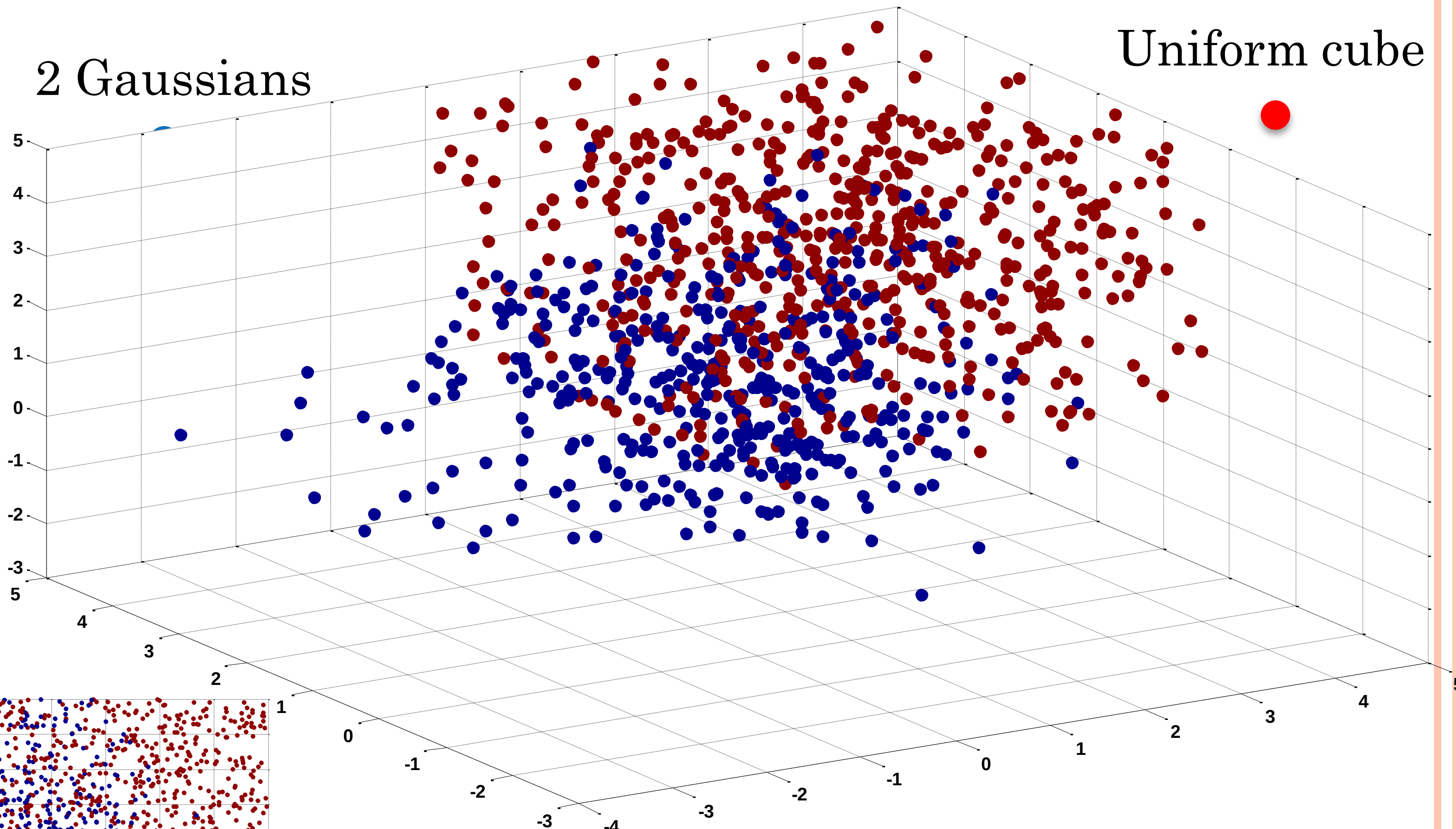
Many users are willing to trade accuracy to better understand the system-yielded results

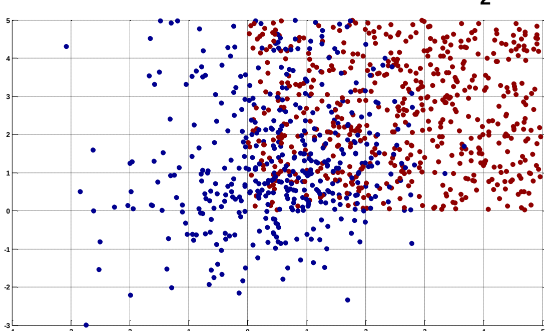
Need: simple, interpretable model

Need: explanatory prediction process

Explanation-Oriented Partitioning (EOP)

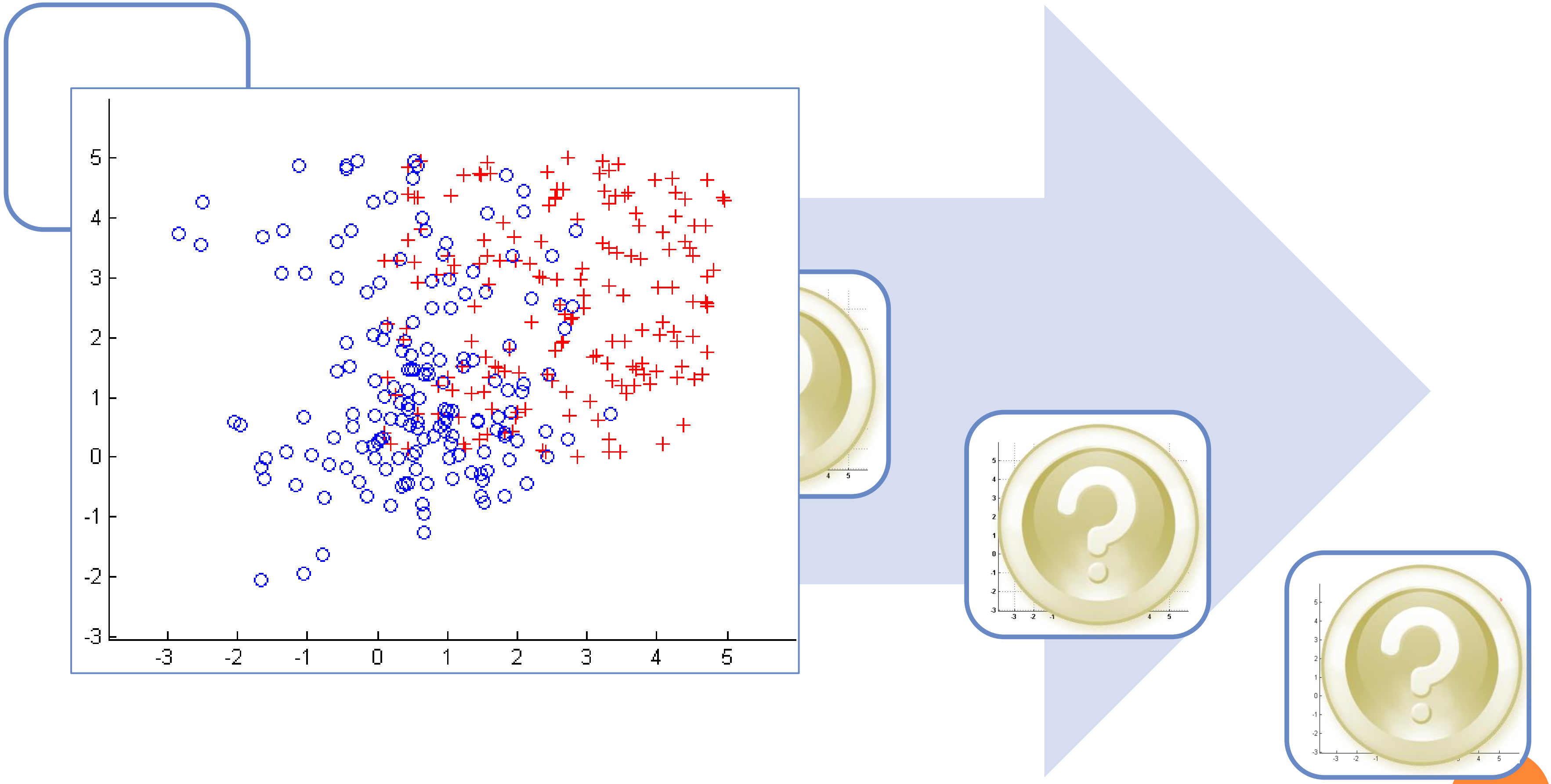
Explanation-Oriented Partitioning (EOP) Execution Example – 3D data





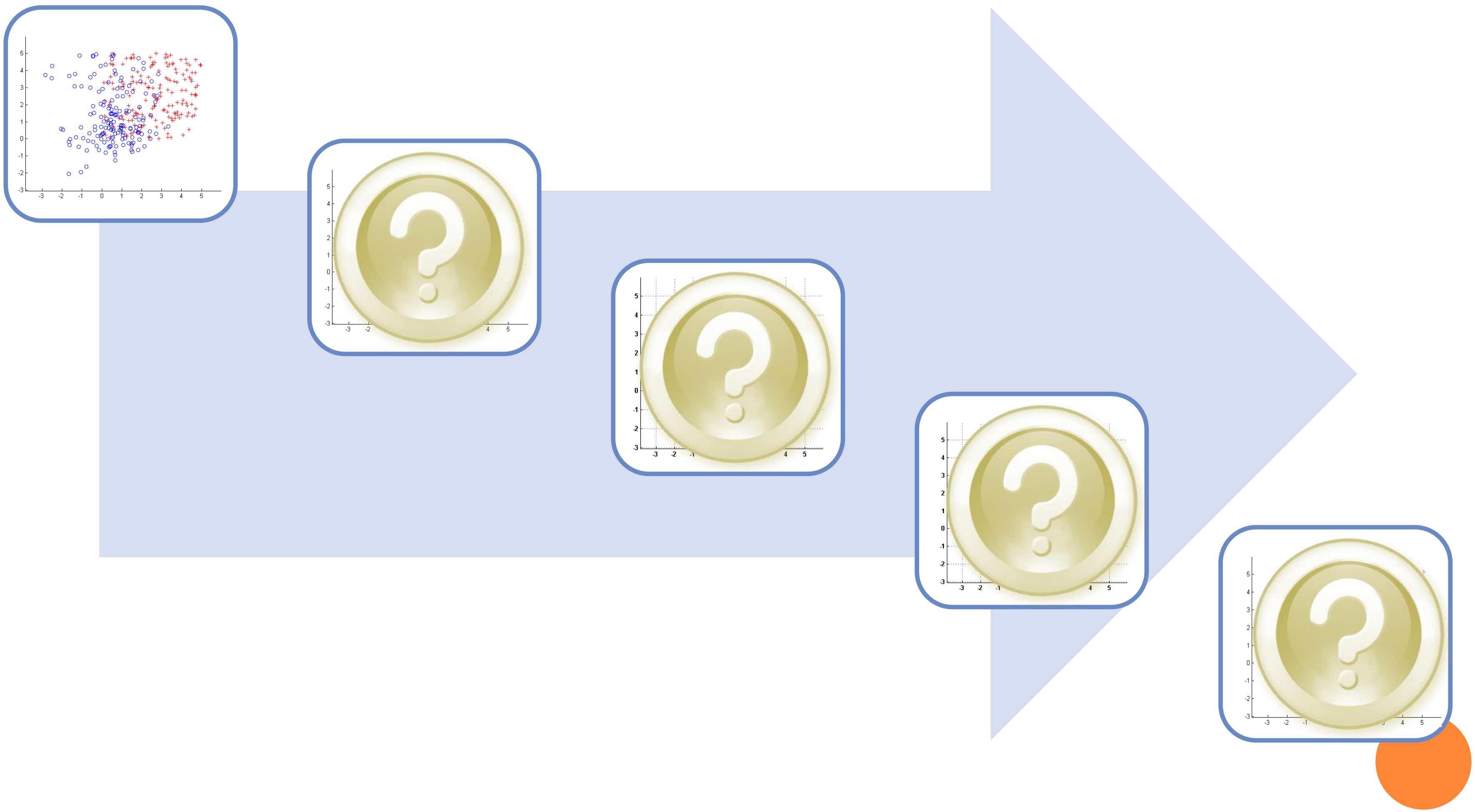
(X,Y) plot

EOP Execution Example – 3D data



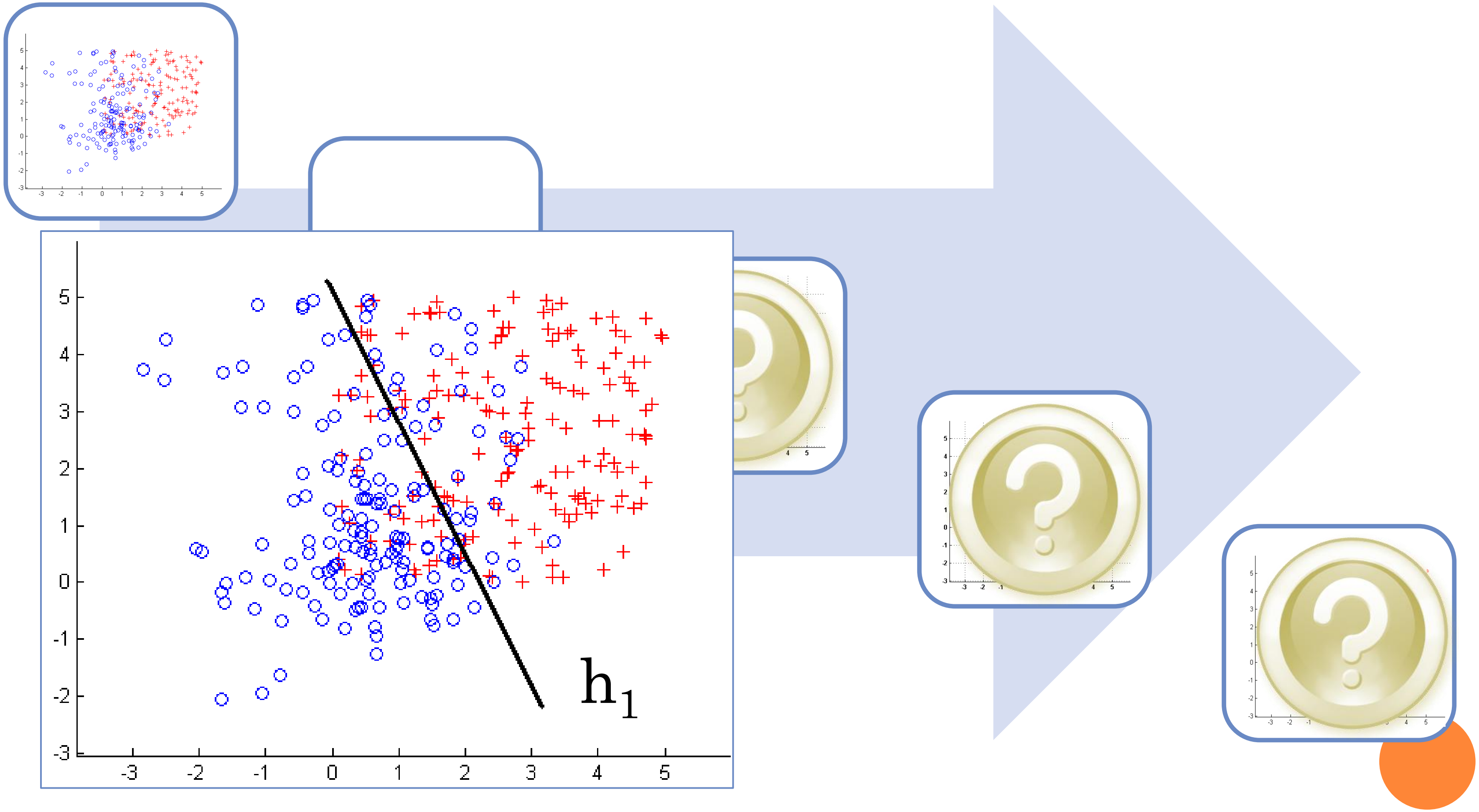
Step 1: Select a projection - (X_1,X_2)

EOP Execution Example – 3D data



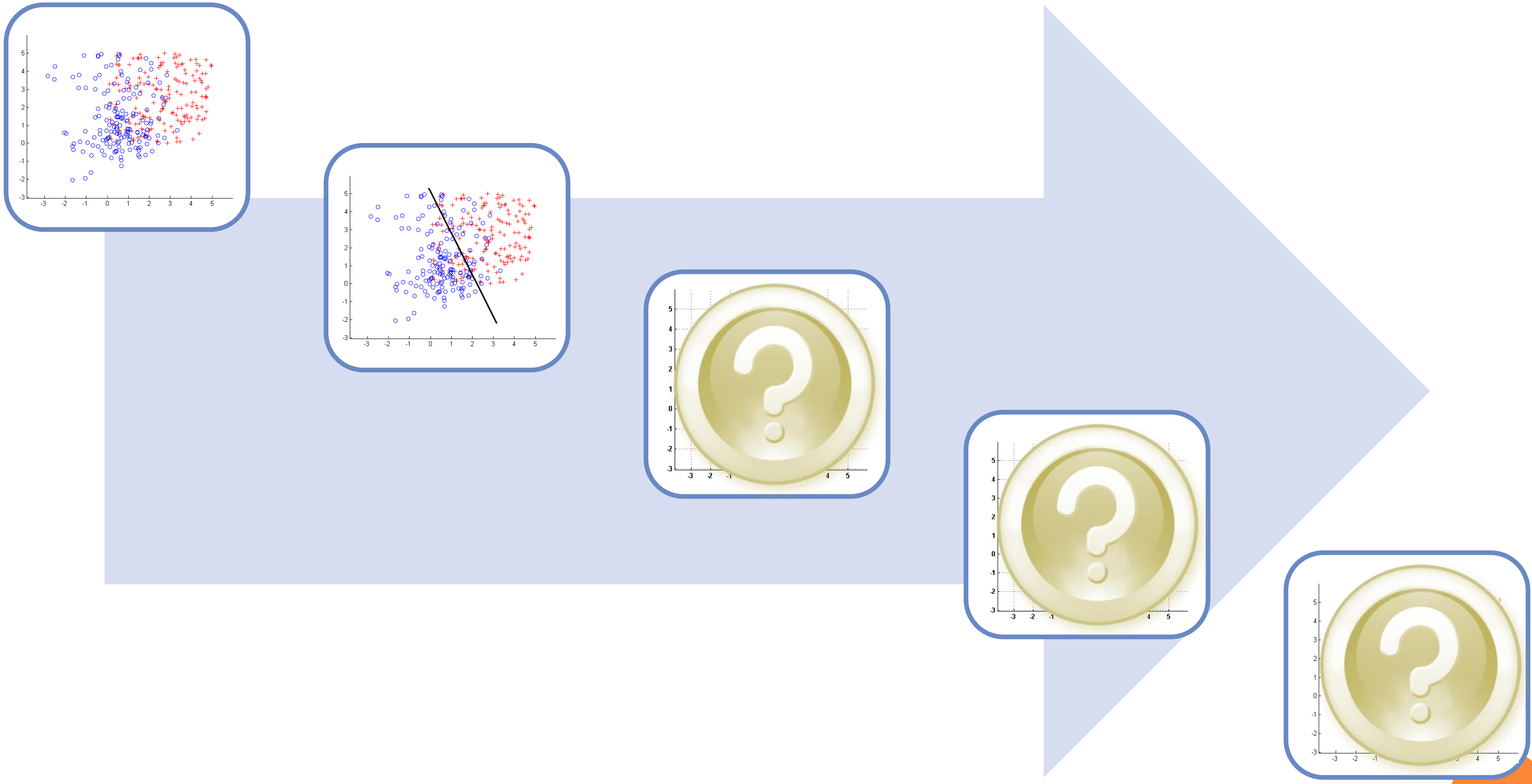
Step 1: Select a projection - (X_1, X_2)

EOP Execution Example – 3D data



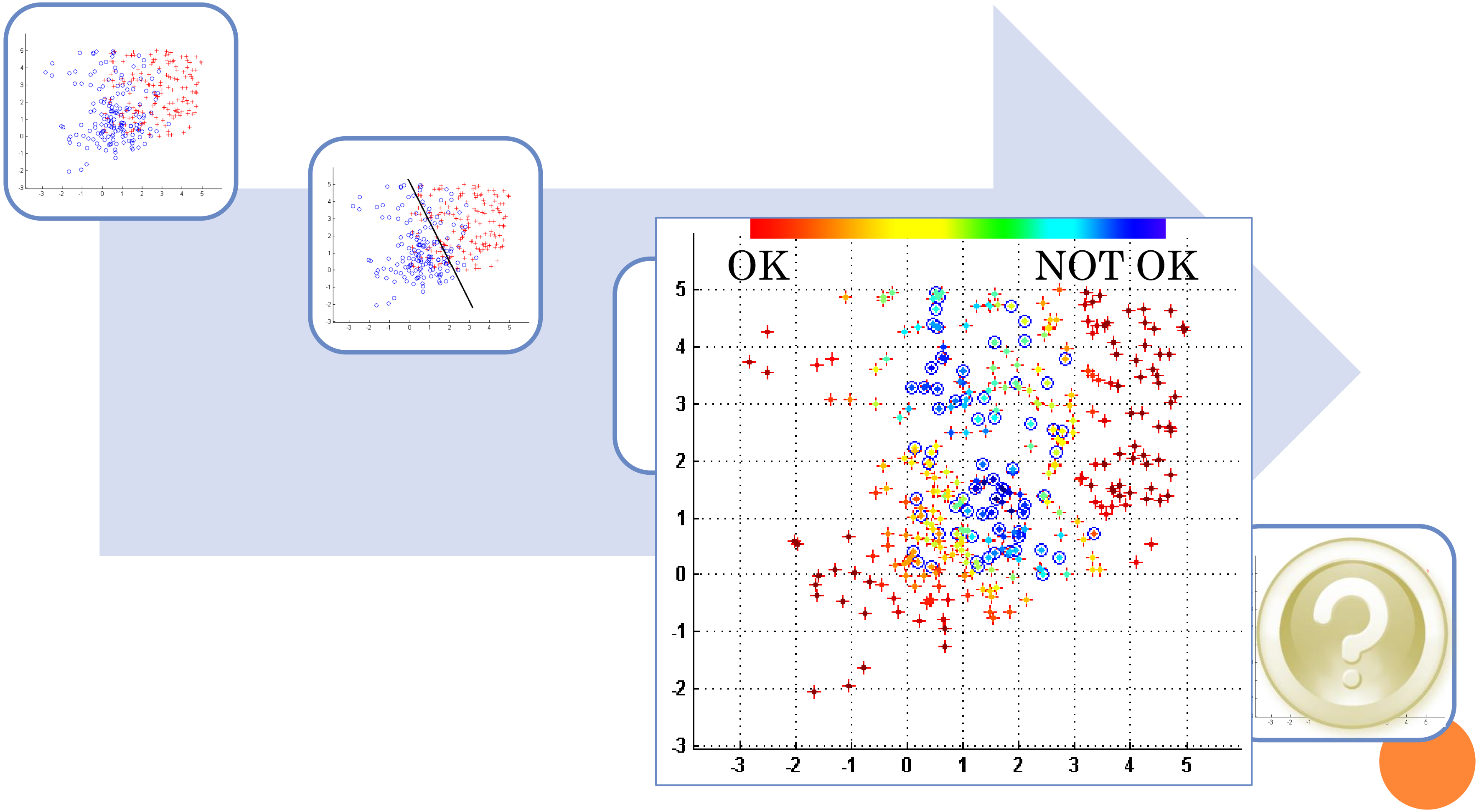
Step 2: Choose a good classifier - call it h_1

EOP Execution Example – 3D data



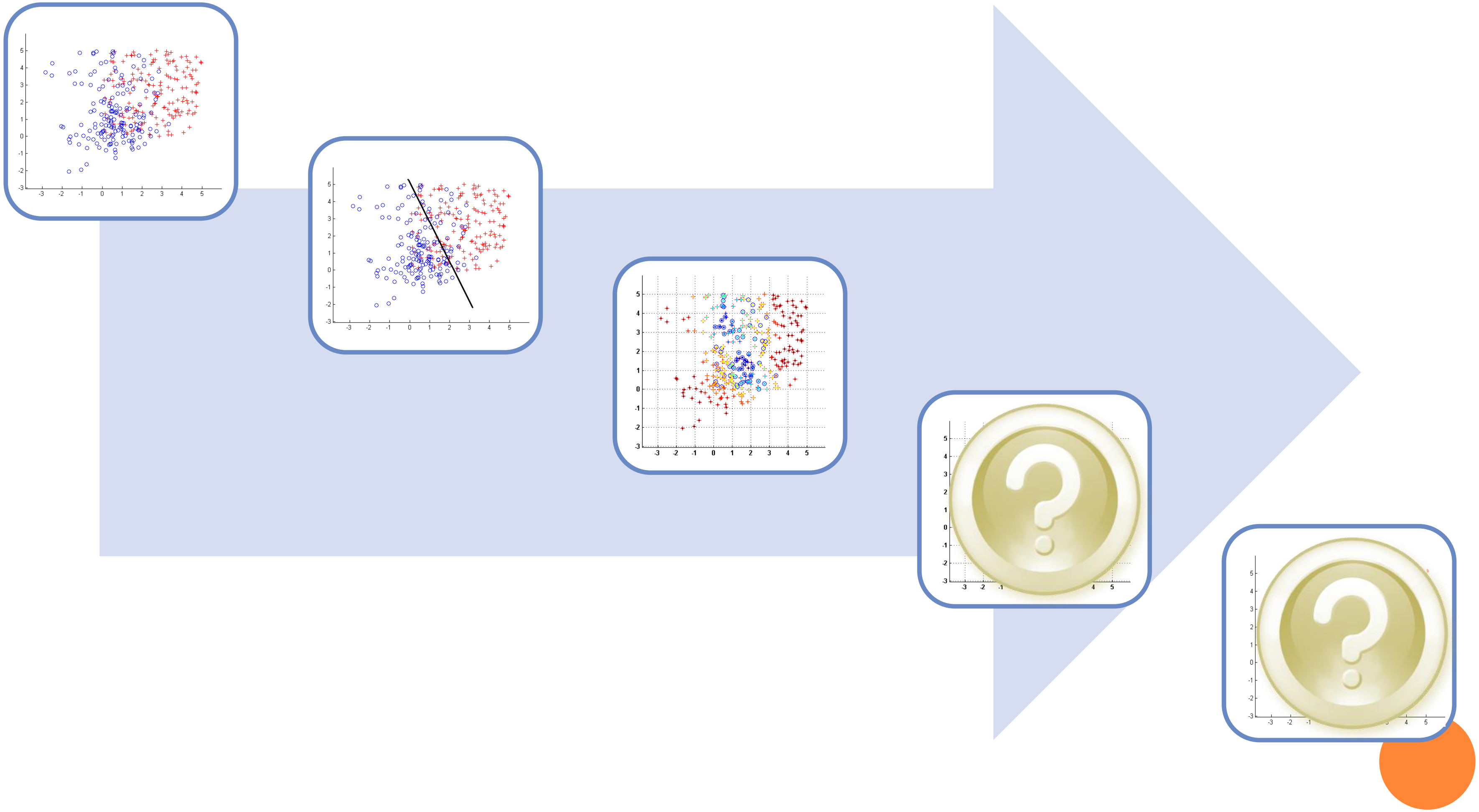
Step 2: Choose a good classifier - call it h_1

EOP Execution Example – 3D data



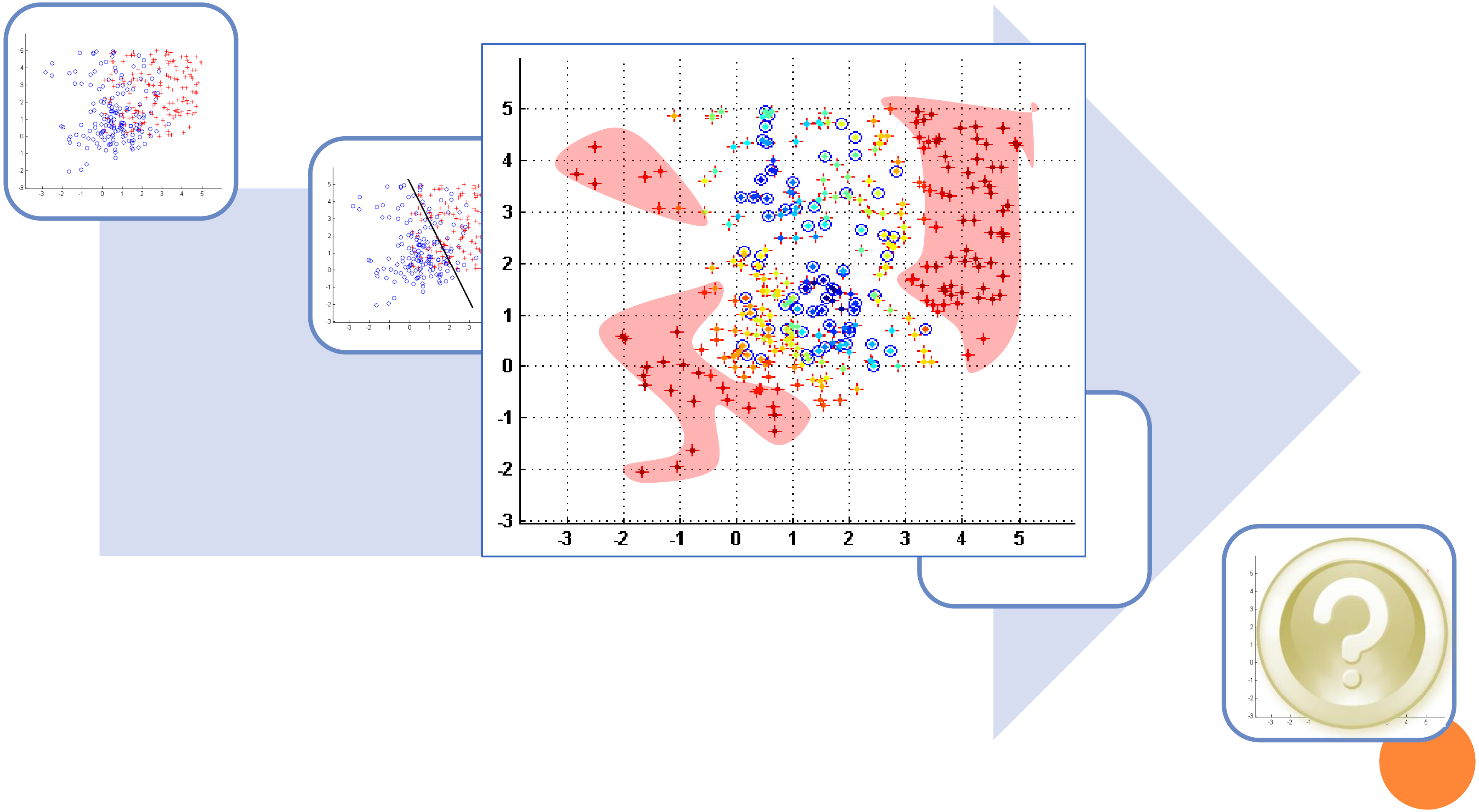
Step 3: Estimate accuracy of h_1 at each point

EOP Execution Example – 3D data



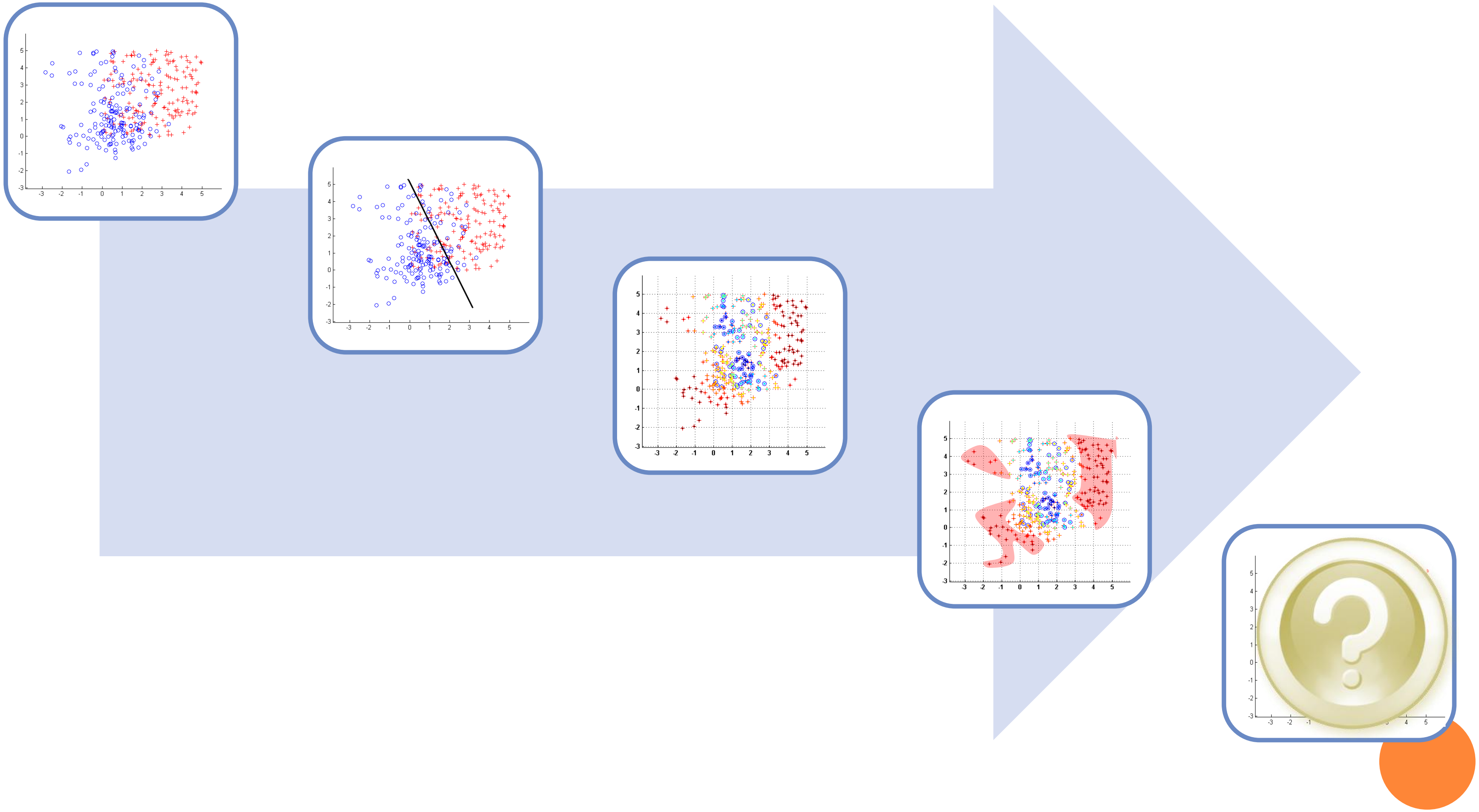
Step 3: Estimate accuracy of h_1 for each point

EOP Execution Example – 3D data



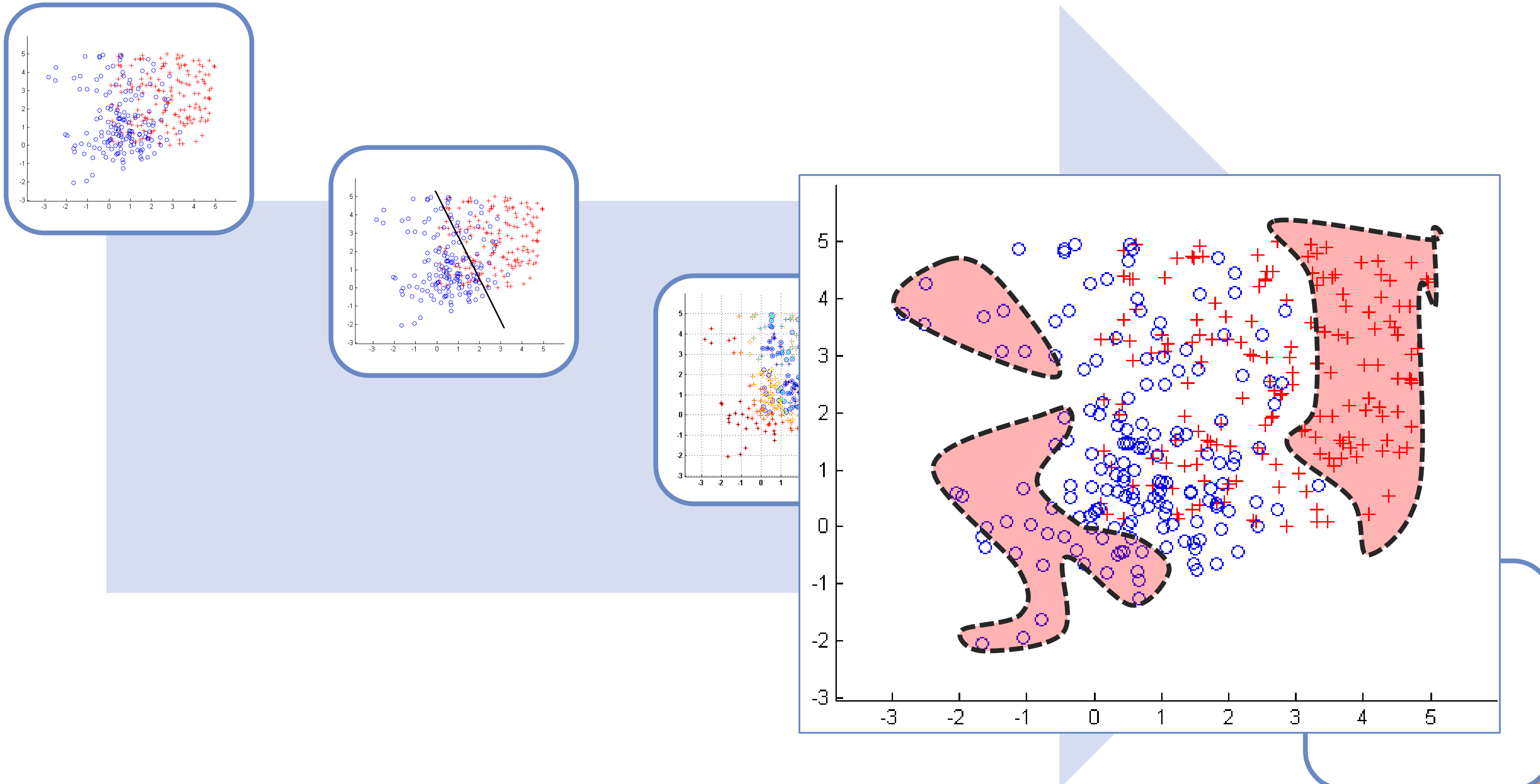
Step 4: Identify high accuracy regions

EOP Execution Example – 3D data



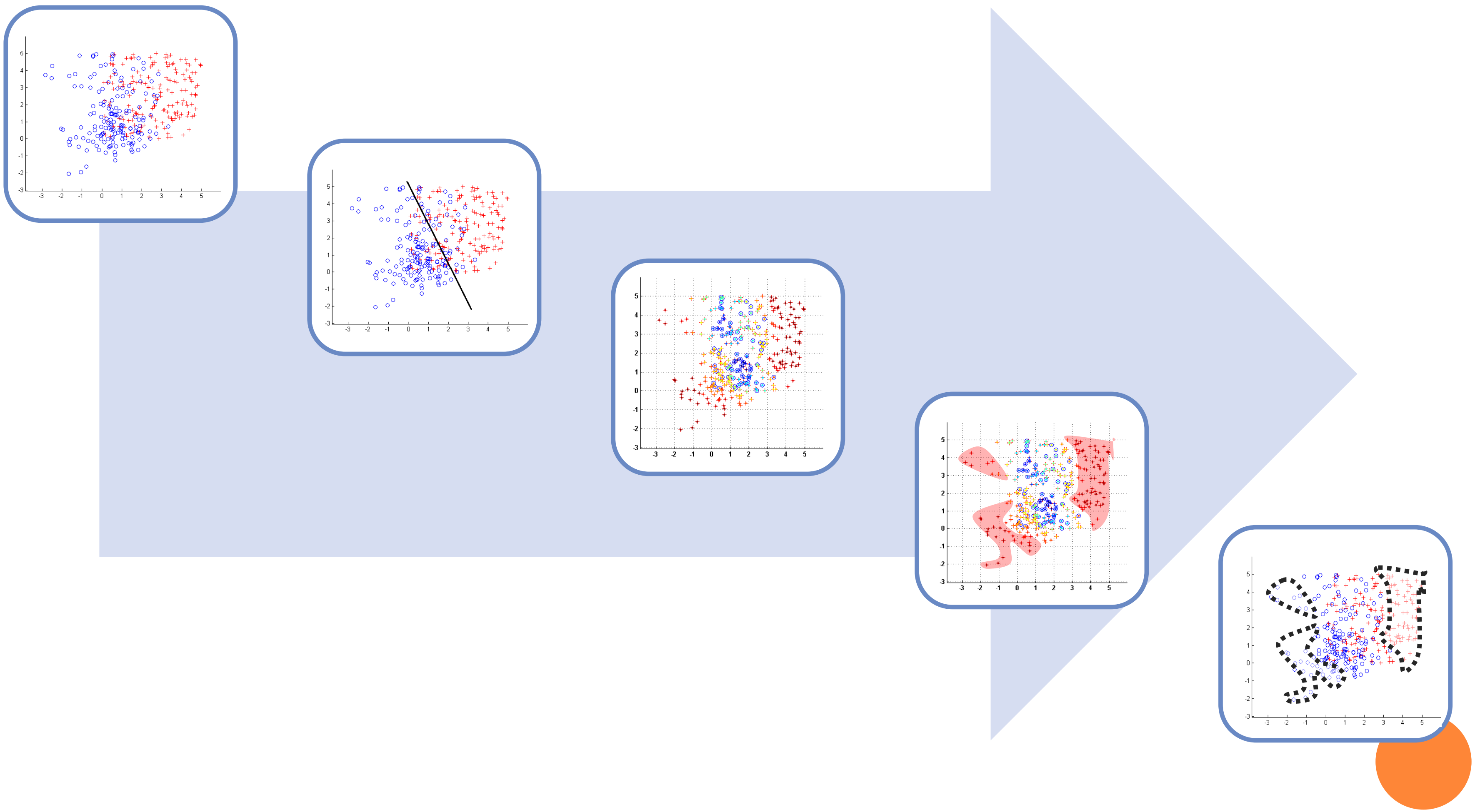
Step 4: Identify high accuracy regions

EOP Execution Example – 3D data



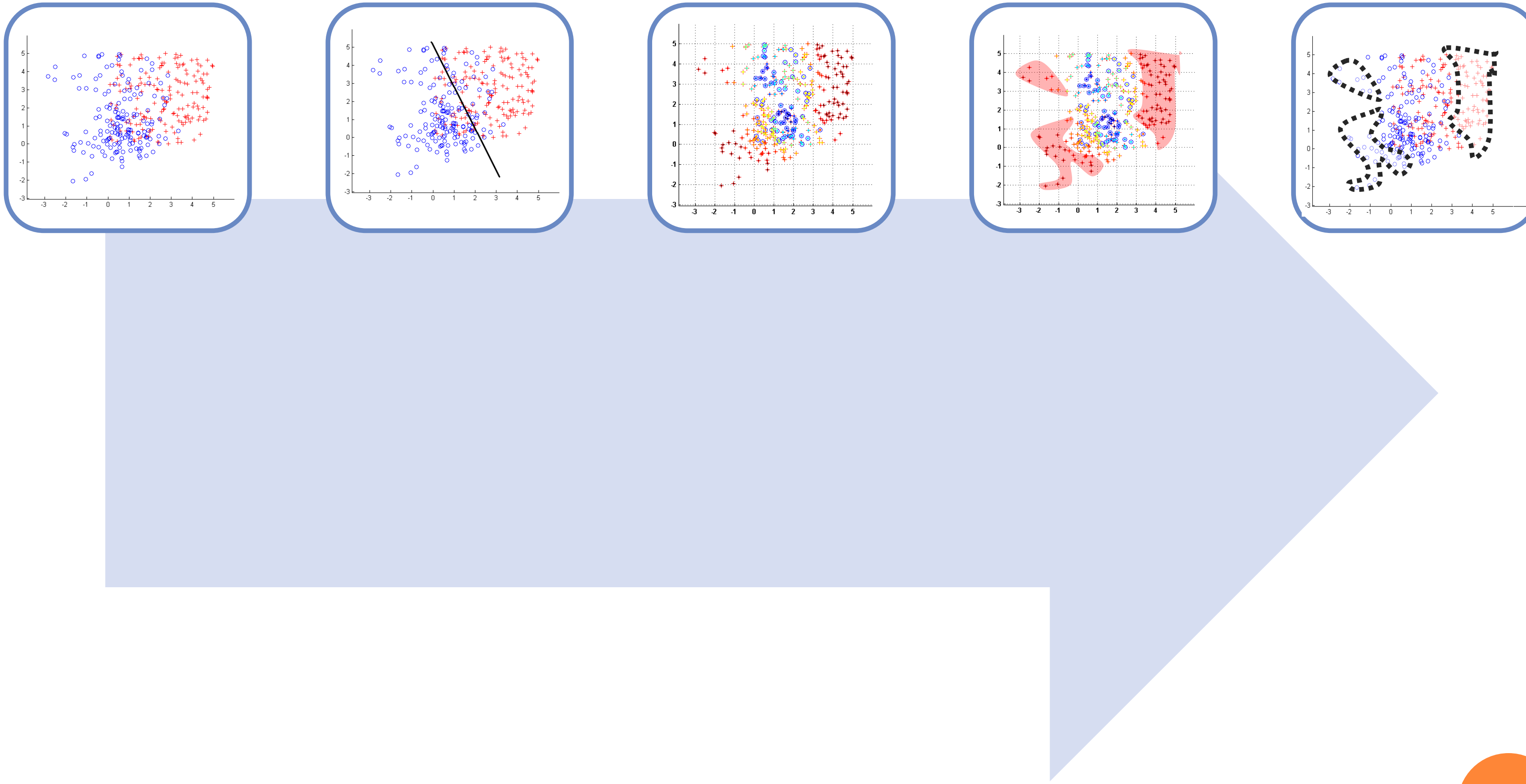
Step 5:Training points - removed from consideration

EOP Execution Example – 3D data



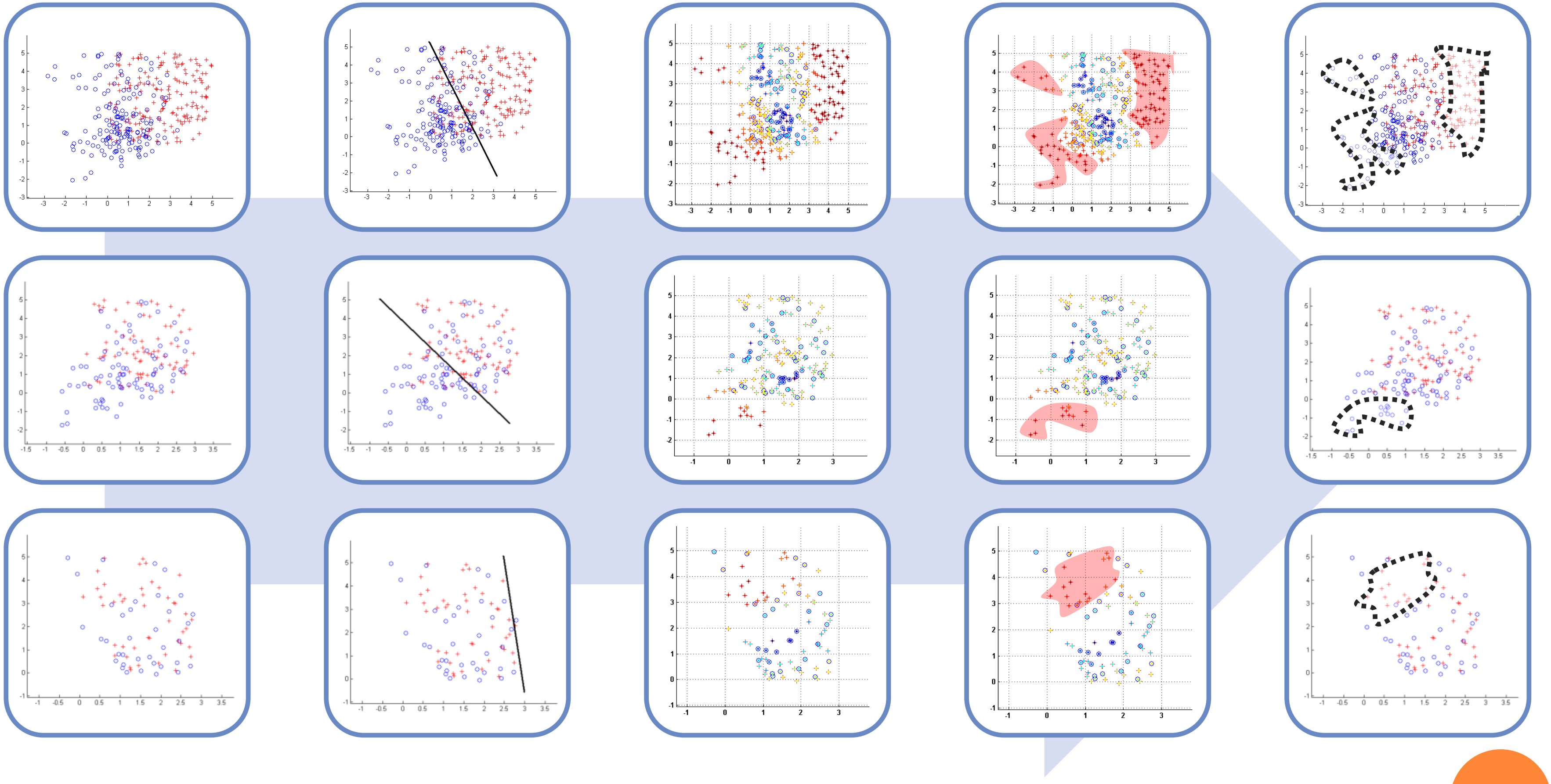
Step 5:Training points - removed from consideration

EOP Execution Example – 3D data



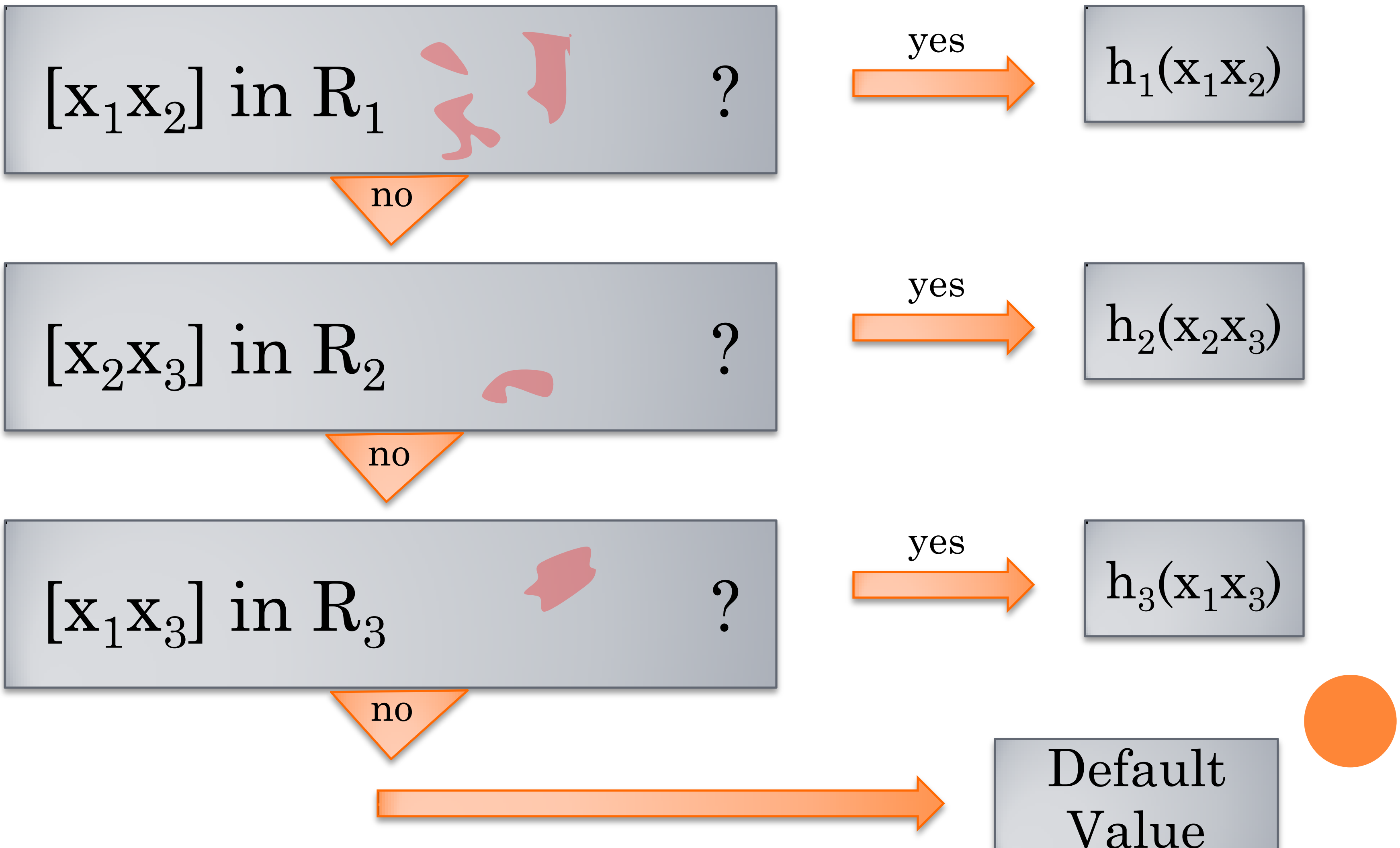
Finished first iteration

EOP Execution Example – 3D data



Iterate until all data is accounted for
or error cannot be decreased

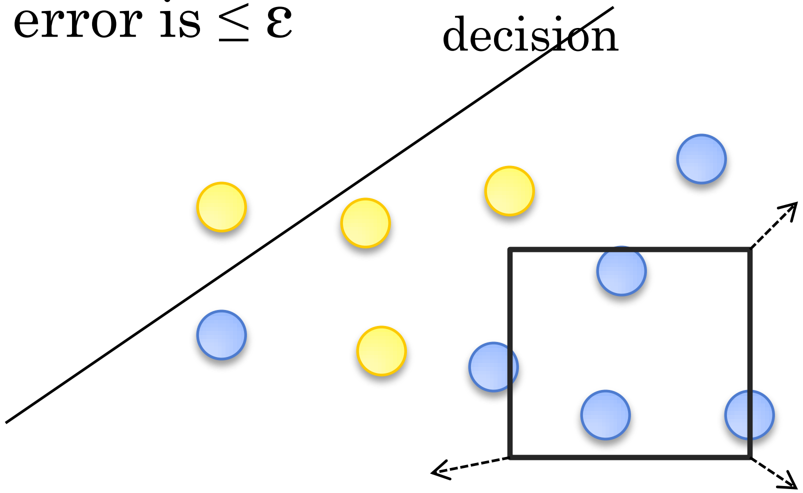
Learned Model – Processing query $[x_1 x_2 x_3]$



PARAMETRIC REGIONS OF HIGH CONFIDENCE (BOUNDING POLYHEDRA)

- Enclose points in simple convex shapes (multiple per iteration)

  Grow contour while train error is ≤ ε



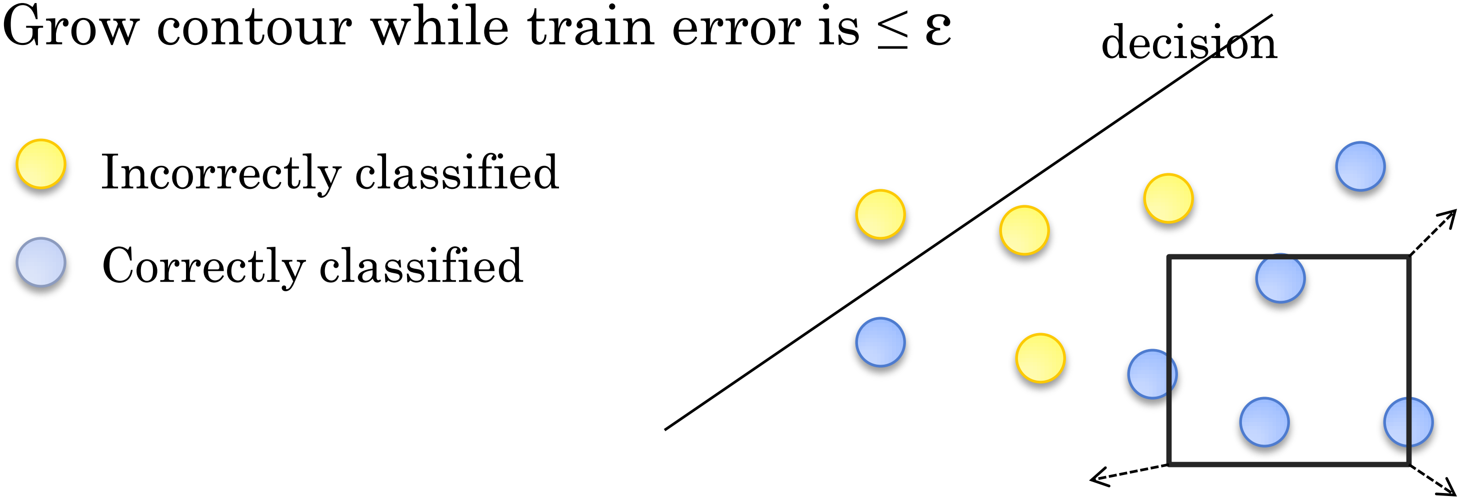
Incorrectly classified

Correctly classified

PARAMETRIC REGIONS OF HIGH CONFIDENCE (BOUNDING POLYHEDRA)

- Enclose points in simple convex shapes (multiple per iteration)

  Grow contour while train error is $\leq \varepsilon$

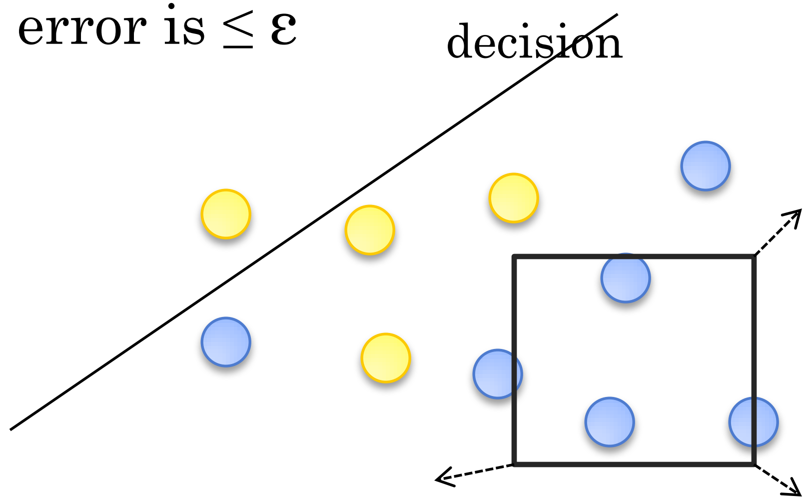


- Calibration on hold out set - remove shapes that:
  - do not contain calibration points
  - over which the classifier is not accurate

Parametric Regions of high confidence (Bounding Polyhedra)

- Enclose points in simple convex shapes (multiple per iteration)

  Grow contour while train error is $\leq \varepsilon$



Incorrectly classified

Correctly classified

- Calibration on hold out set - remove shapes that:
  - do not contain calibration points
  - over which the classifier is not accurate

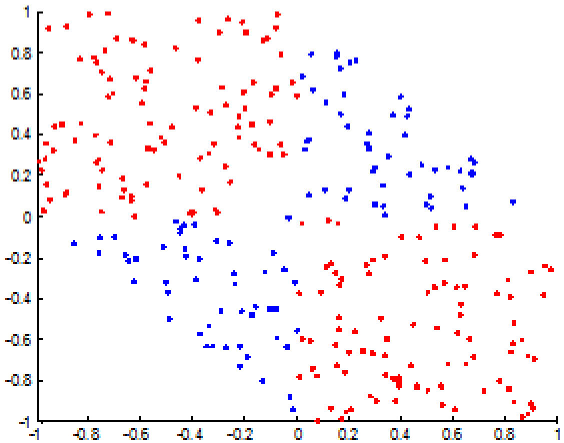
👍 Intuitive, visually appealing - hyper-rectangles/spheres

Outline

- Motivation of need for interpretability
- Explanation-Oriented Partitioning (EOP)
- Evaluation of EOP
- Summary

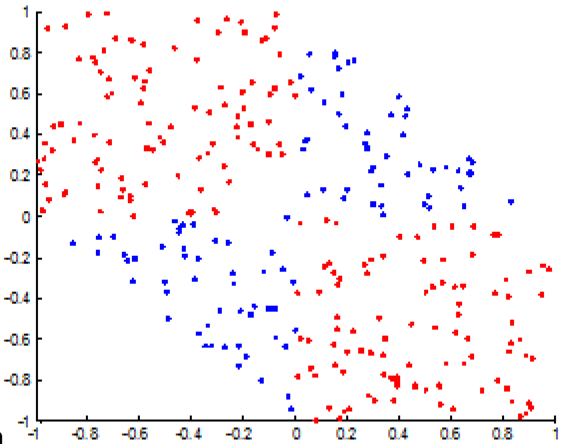
Benefits of EOP - Avoiding Needless Complexity -

Typical XOR dataset

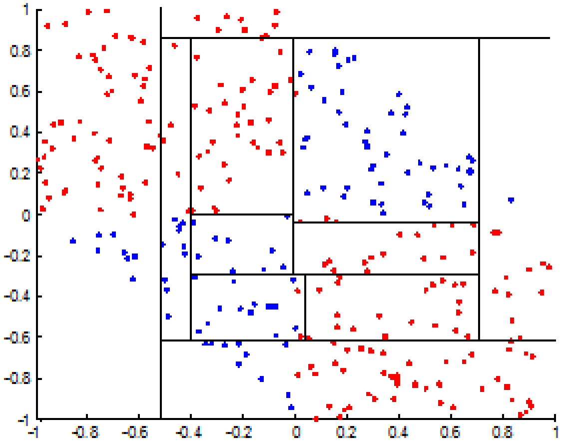


Benefits of EOP - Avoiding Needless Complexity -

Typical XOR dataset

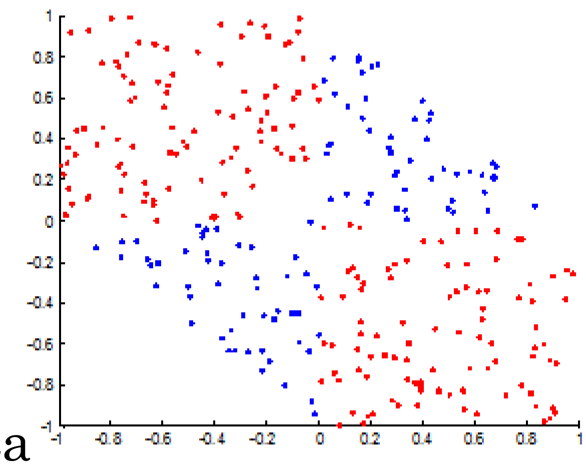


CART
- is accurate
- takes many iterations
- does not uncover or leverage structure of data

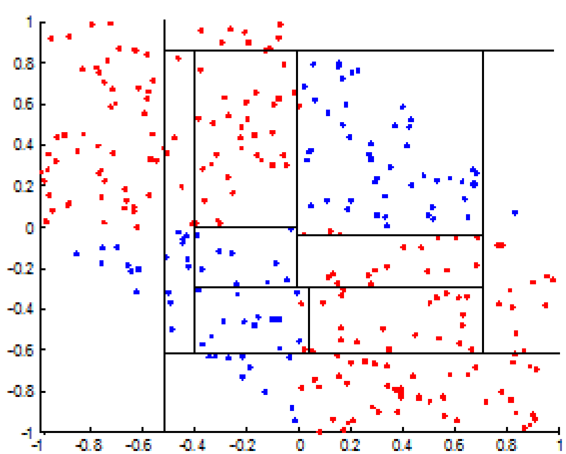


Benefits of EOP
- Avoiding Needless Complexity -

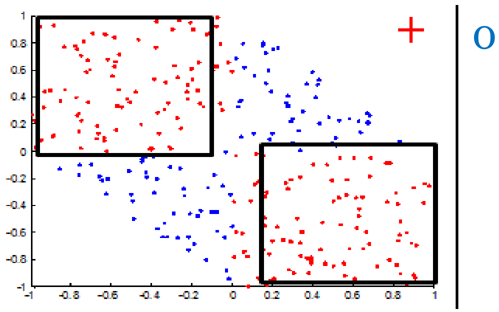
Typical XOR dataset



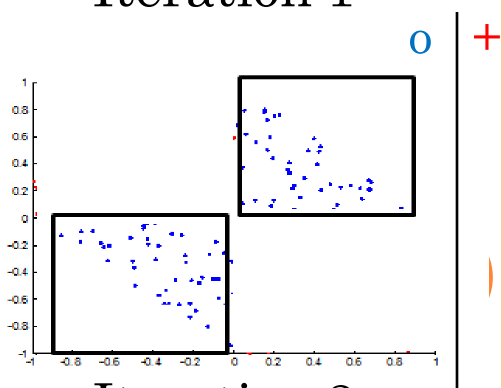
CART
- is accurate
- takes many iterations
- does not uncover or leverage structure of data



EOP
- equally accurate
- uncovers structure



Iteration 1



Iteration 2

Comparison To Boosting

- What is the price of understandability?

- Why boosting?
  - It is an [arguably] good black-box classifier
  - Learns an *ensemble* using any type of classifier
  - Iteratively targets *data misclassified earlier*

- Criterion: *Complexity of the resulting model*

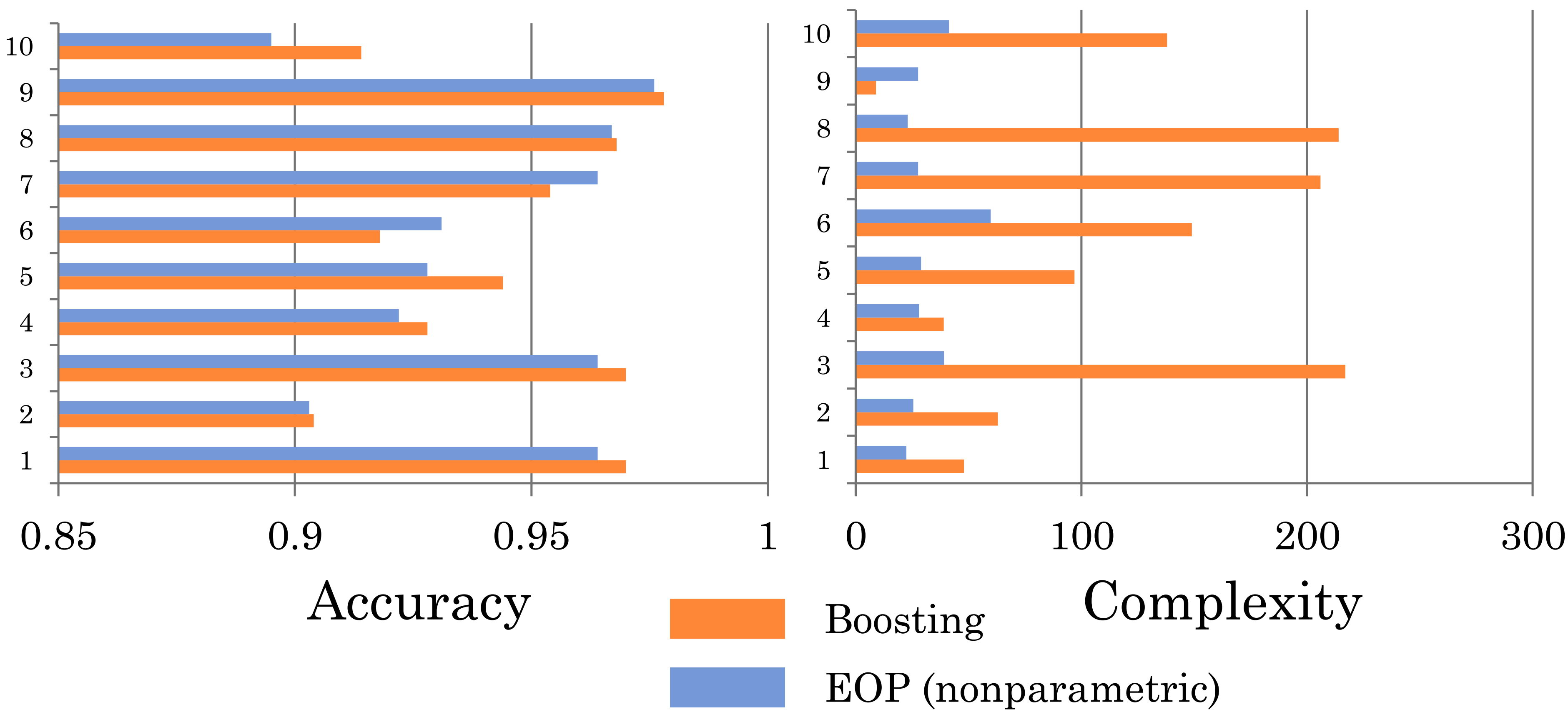
  = number of vector operations to make a prediction

Comparison to boosting - Setup

- Problem: Binary classification
- 10D Gaussians/uniform cubes for each class
- Statistical significance: repeat experiment with several datasets and compute paired t-test p-values
- Results obtained through 5-fold cross validation

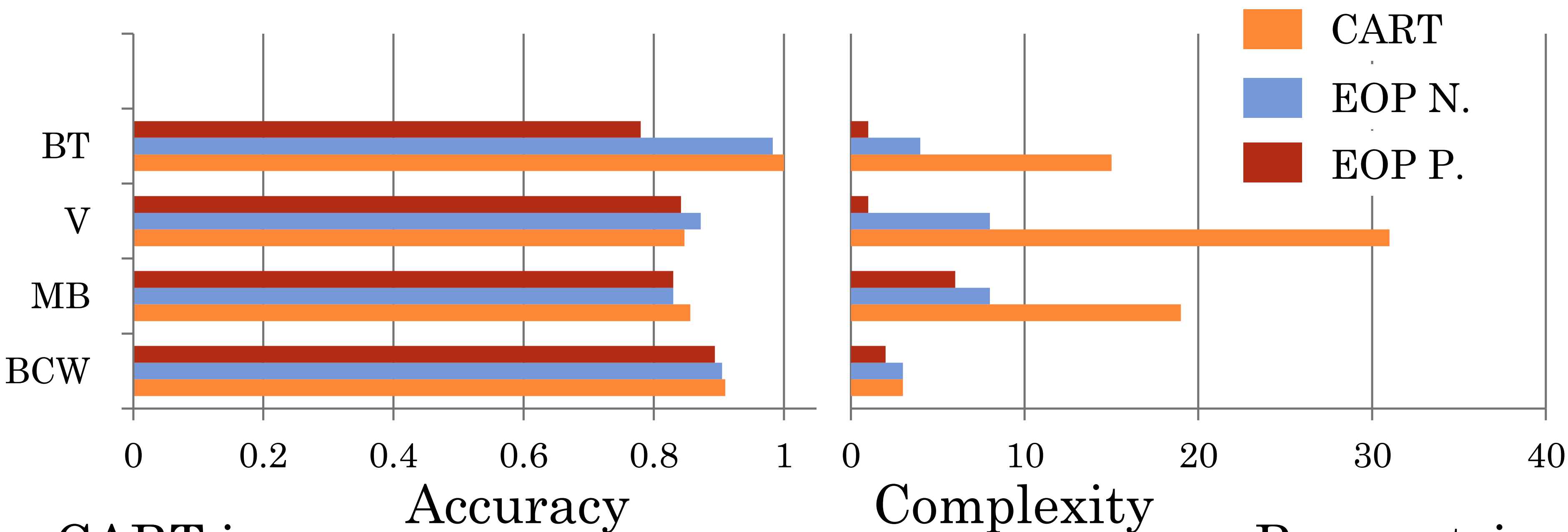
EOP VS ADABOOST - SVM BASE CLASSIFIERS

- EOP is often less accurate, but not significantly
- the reduction of complexity is statistically significant



Accuracy p-value: 0.832 **Complexity p-value: 0.003**

EOP (STUMPS AS BASE CLASSIFIERS) VS CART DATA FROM THE UCI REPOSITORY

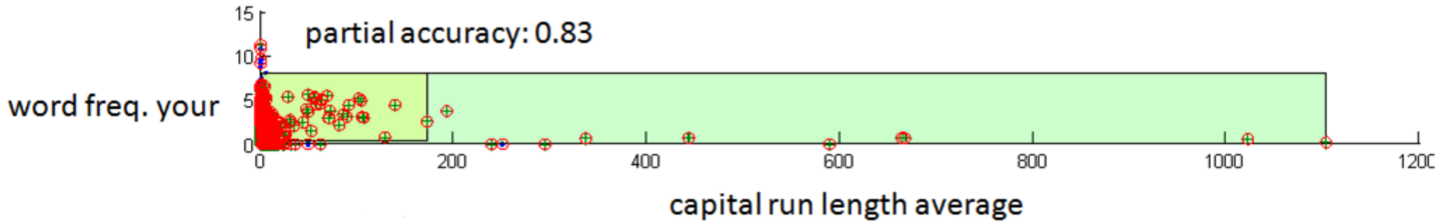


- CART is the most accurate

Dataset	*# of Features*	*# of Points*
Breast Tissue	10	1006
Vowel	9	990
MiniBOONE	10	5000
Breast Cancer	10	596

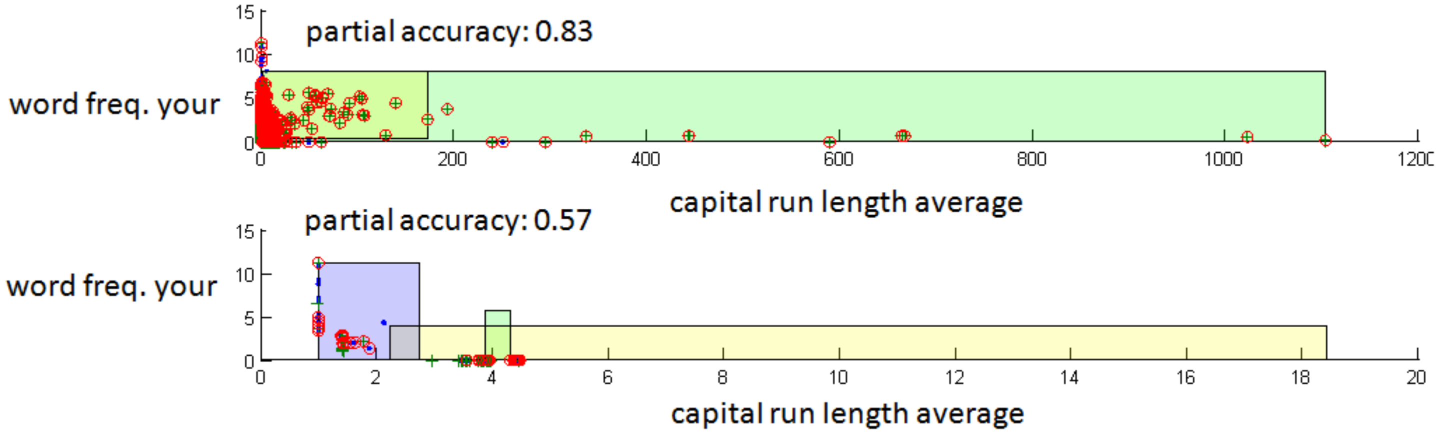
- Parametric EOP yields the simplest models

Explaining Real Data - Spambase



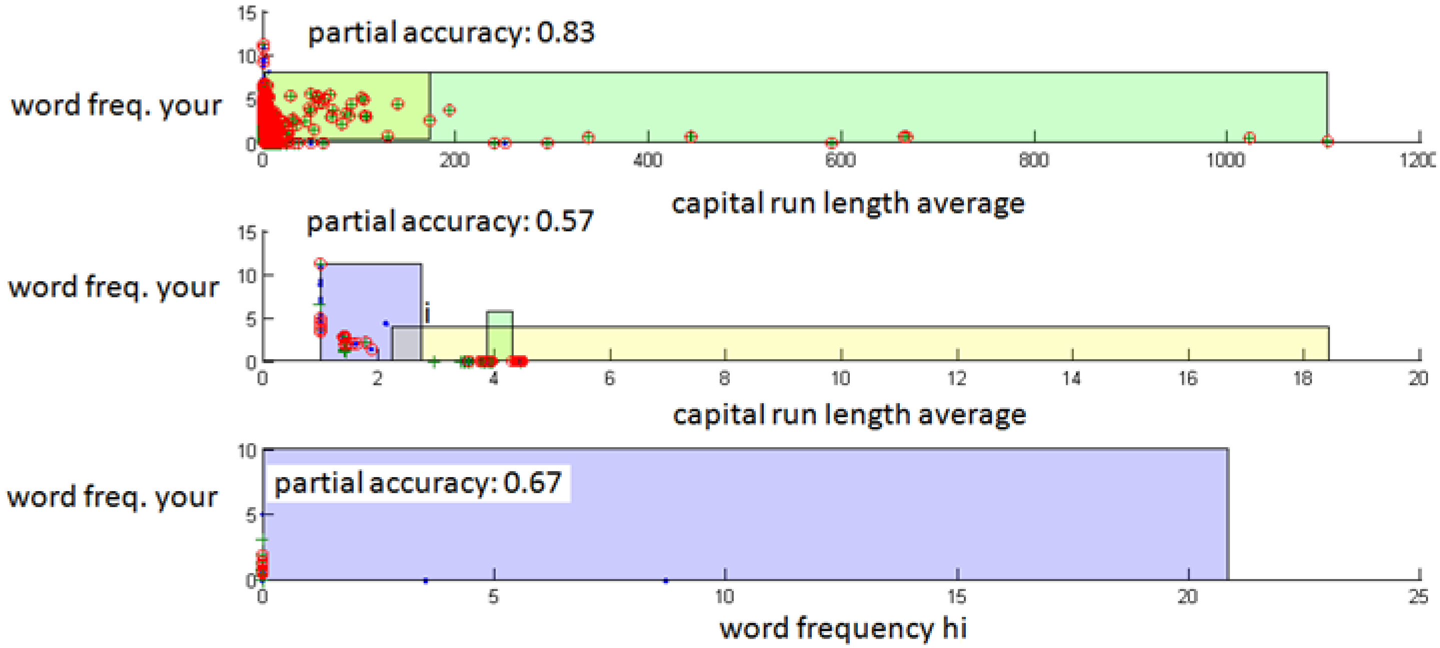
- 1st Iteration
  - classier labels everything as spam
  - high confidence regions do enclose mostly spam and
    - Incidence of the word 'your' is low
    - Length of text in capital letters is high

Explaining Real Data - Spambase



- 2nd Iteration
  - the threshold for the incidence of `your' is lowered
  - the required incidence of capitals is increased
  - the square region on the left also encloses examples that will be marked as `not spam'

Explaining Real Data - Spambase



- 3rd Iteration
  - Classifier marks everything as spam
  - Frequency of 'your' and 'hi' determine the regions

SUMMARY

- EOP maintains classification accuracy but uses *less complex models* when compared to Boosting

- EOP with decision stumps finds *less complex models* than CART at the price of a small decrease in accuracy

- EOP gives interpretable high accuracy regions

- We are currently testing EOP in a range of practical application scenarios

Thank You

Extra Results

Explaining real data - fuel