Applied Statistics for Neuroscientists
Part IIa: Machine Learning

Dr. Seyed-Ahmad Ahmadi

16.11.2017

Outline – Machine Learning

- “Difference” between statistics and machine learning
- Modeling the problem with (un-)supervised learning
- Typical Machine Learning pipeline
- Decision functions
- Important supervised learning algorithms:
  - kNN / SVM / Decision trees → Random Forests
- Cross-validation
- Concrete example from my research
  - Anatomy locatlization in images with RF regression
- Outlook:
  - Clustering
  - Manifold learning
- Questions?

Statistics vs. machine learning

According to Larry Wassermann
(professor for dept.s for statistics and Machine Learning at Carnegie Mellon University)

> *"The short answer is: None. They are both concerned with the same question: how do we learn from data?"*

- Statistics emphasizes **formal statistical inference** (confidence intervals, hypothesis tests, optimal estimators) in **low dimensional** problems.
  - Survival analysis, spatial analysis, multiple testing, minimax theory, deconvolution, semiparametric inference, bootstrapping, time series
- Machine Learning emphasizes **high dimensional prediction** problems.
  - Online learning, (semi-)supervised learning, manifold learning, active learning, boosting

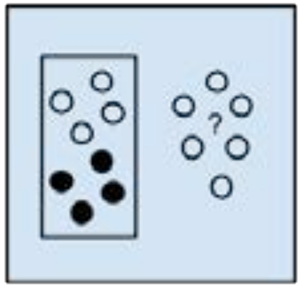
Mindmap of ML algorithms

Adapted from: https://machinelearningmastery.com

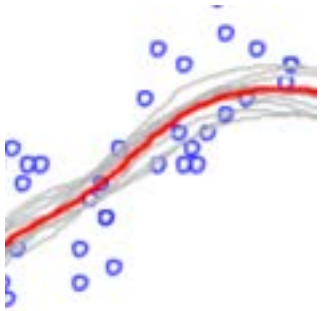
ML categories: Modeling your problem

Supervised Learning

Classification

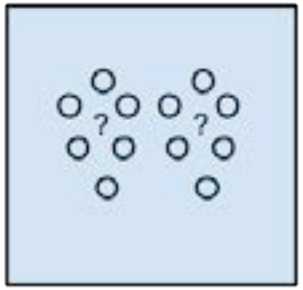


Regression

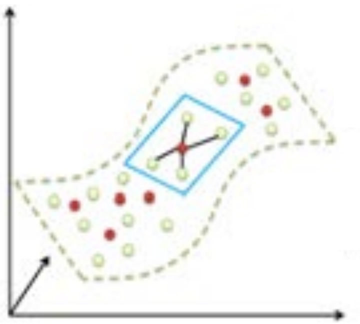


Unsupervised Learning

Clustering



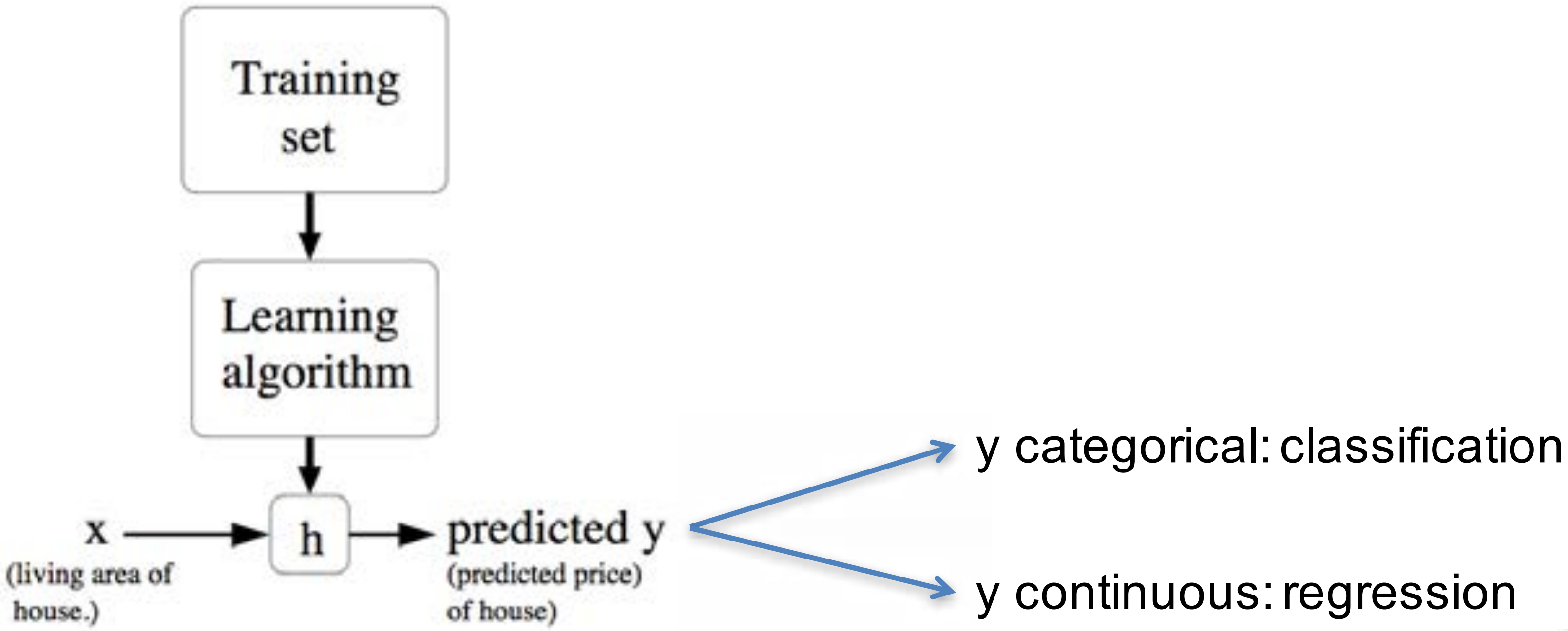
Manifold learning



Adapted from: https://machinelearningmastery.com

Semi-supervised / Reinforcement Learning

Supervised Learning: Classification and regression

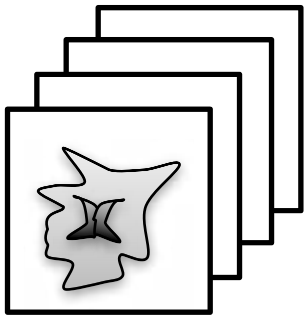


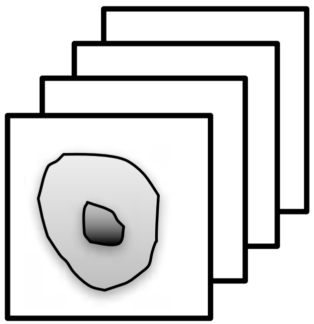
Adapted from: Andrew Ng, CS 229, fall 2012, Stanford university

A typical ML pipeline

Data acquisition: **Pre-processing:** **Feature extraction:** **Machine learning:**

Cancer cells

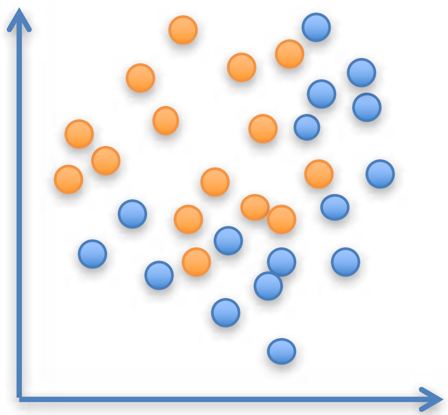




Healthy cells

Image normalization
De-noising
Cell localization
ROI cropping
...

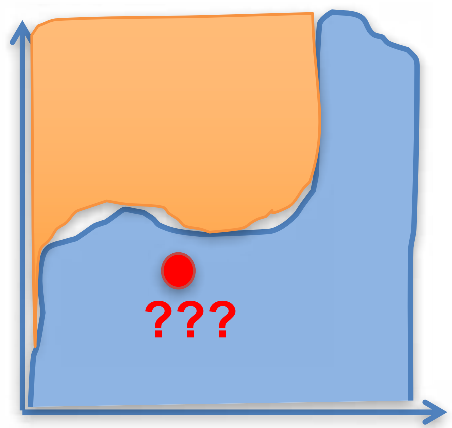
„*Mitosis grade*“



„*Cell roundness*“

Features: $x \in X \subset R^n$

Classes: $c \in C \cup [0,1]$



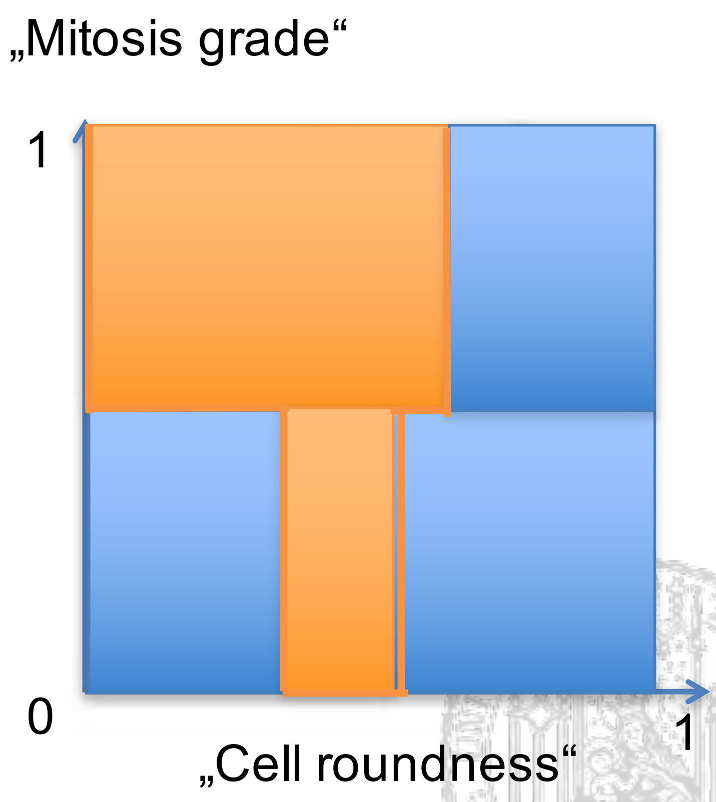
Inference for previously unseen data points

$$c = f(x)$$

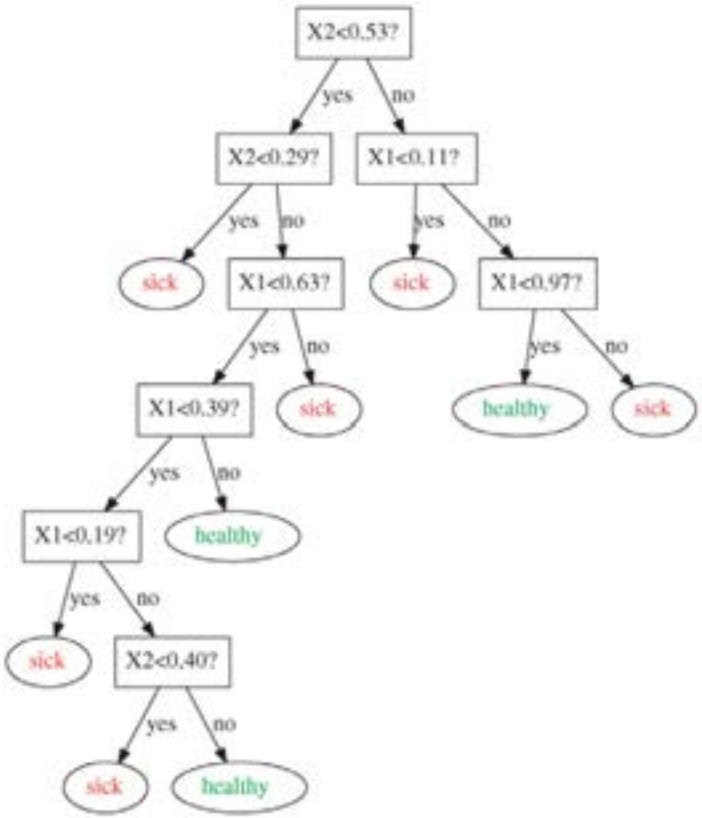
Decision functions – „hand-made“

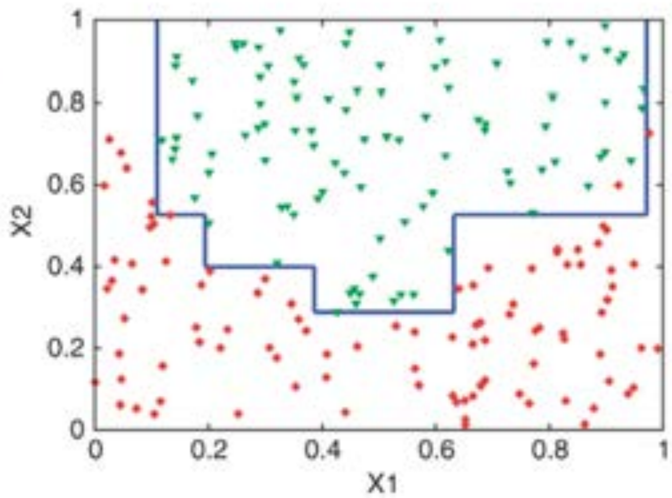
Rule-based algorithm / Decision tree

- If mitosis > 0.5: cancer
  - If roundness >0.6: healthy
- Else
  - If 0.4 < roundness < 0.5: cancer



Decision trees

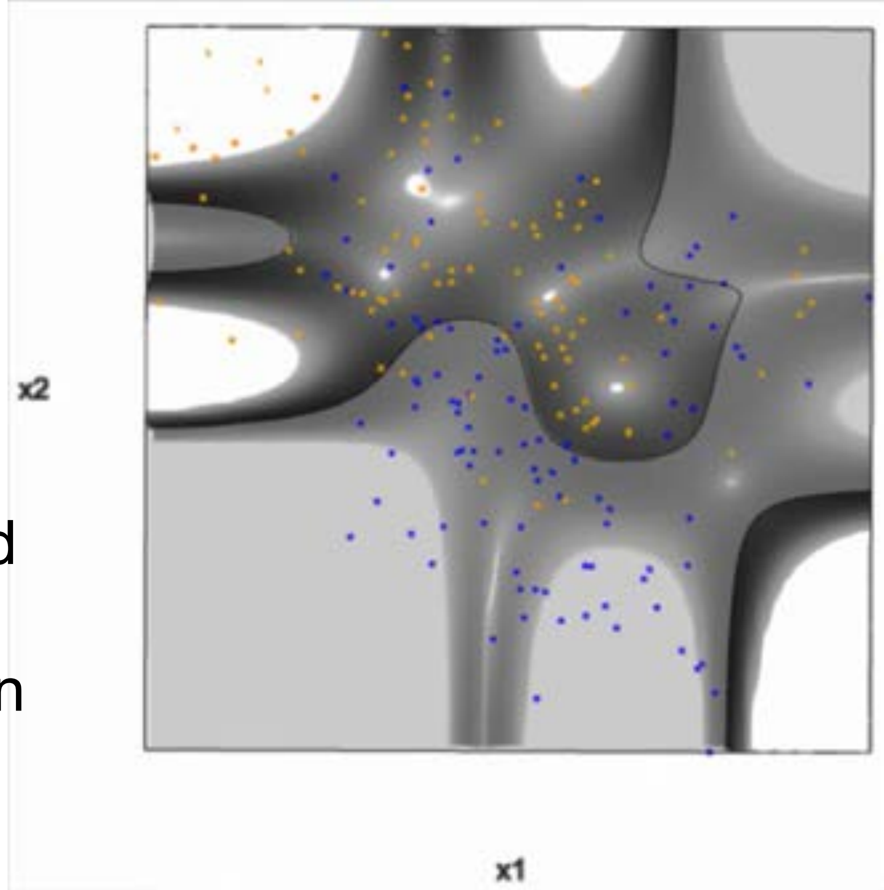




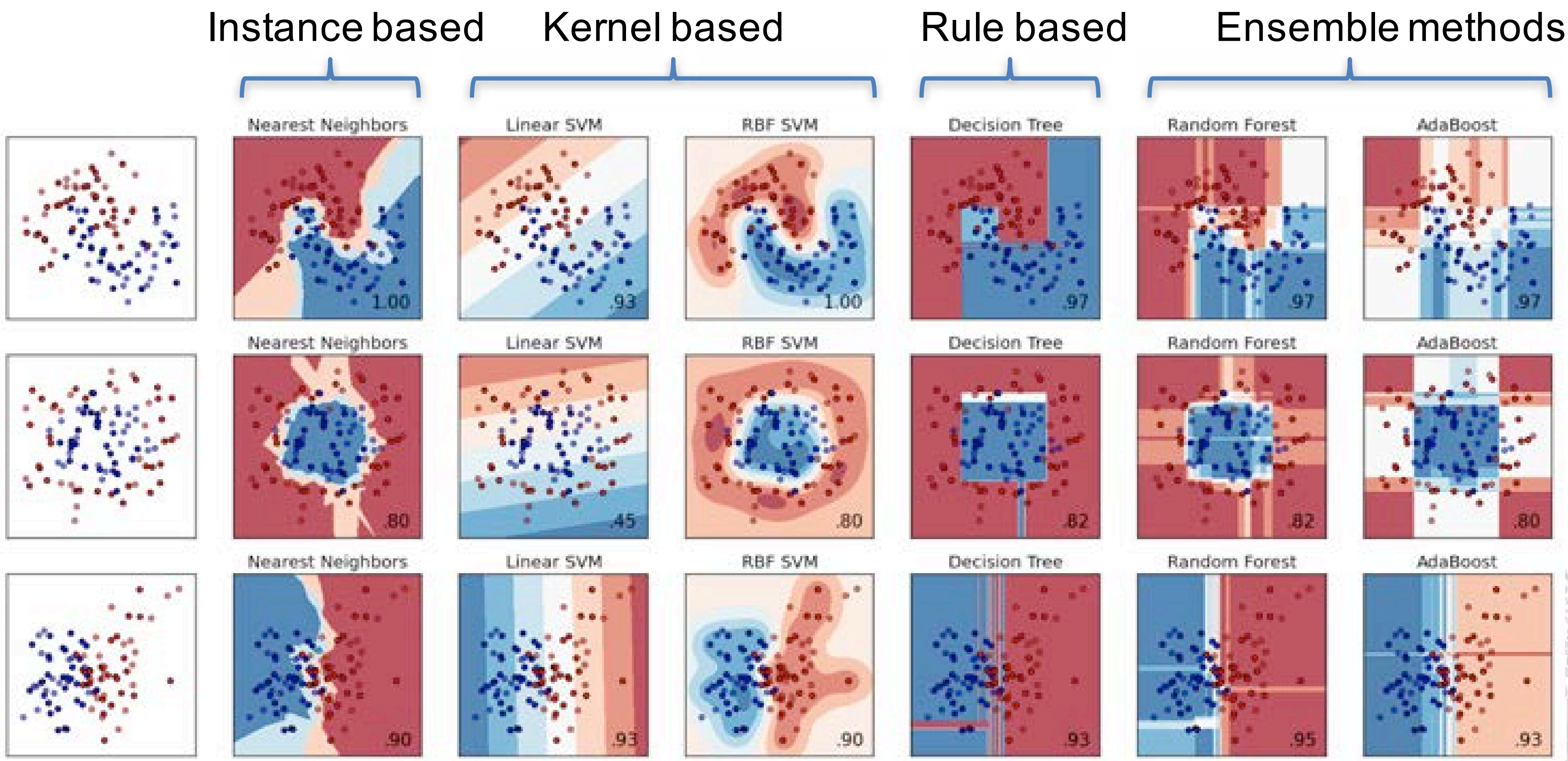
P. Geurts et al., Supervised learning with decision tree-based methods in computational and systems biology, Mol. BioSyst. 5 (2009) 1593–1605

Decision functions

- A ML algorithm tries to automatically find an optimal decision boundary to separate classes
- The decision boundary is the threshold of a decision function
- A ML algorithm fits the decision function to the data in order to find the optimal boundary
- Devision boundary (on the right) in 3D: http://biostatmatt.com/HTF-Figure-5.11b/
- NB: The decision boundary is:
  - Maximally separating
  - Highly non-linear



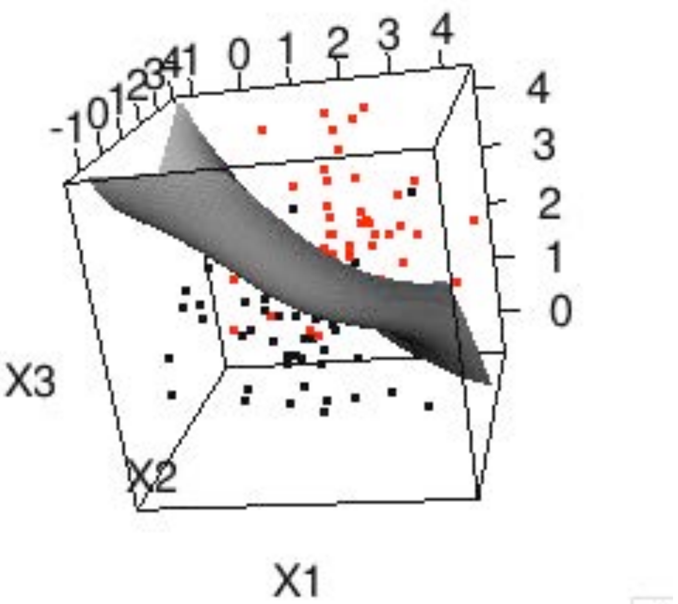
Different classifiers and their decision boundaries



Adapted from: http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

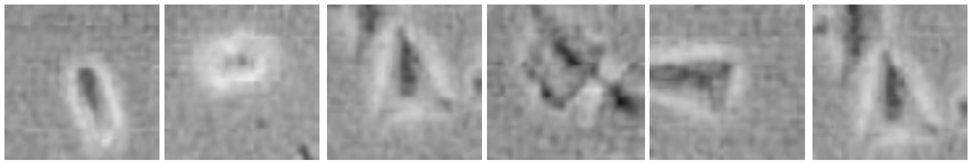
Machine learning helps with high-dimensional, complex problems

Decision function visualized for
3-dimensional feature space



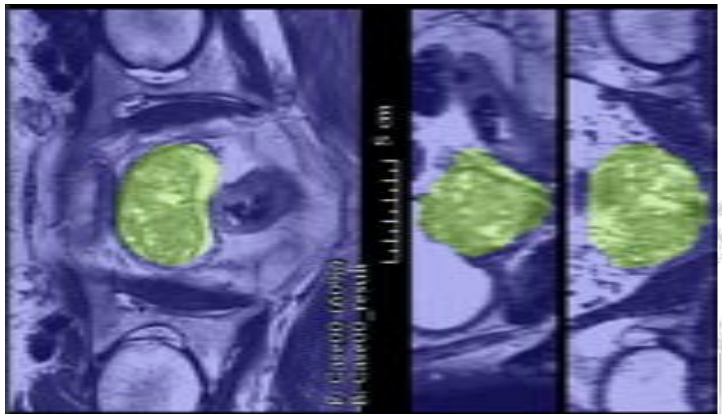
http://stackoverflow.com/questions/10266195/plot-svm-in-3-dimension

Classification of 32 x 32 cell images:
1024 dimensional feature space



e.g. Javidi, Image Recognition and Classification: Algorithms, Systems, and Applications, p. 462, CRC press, 2002

Segmentation of 128 x 128 x 64 MRI volumes:
1.048.576 dimensional feature space

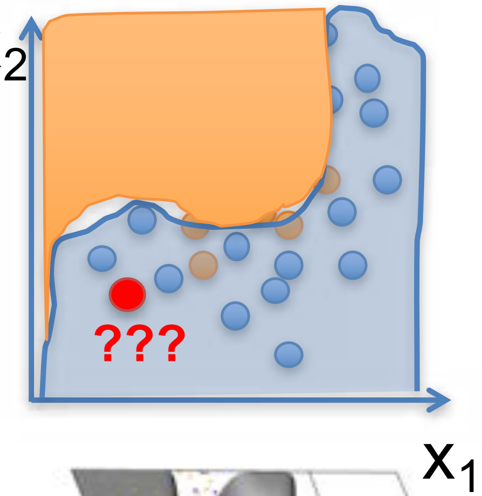


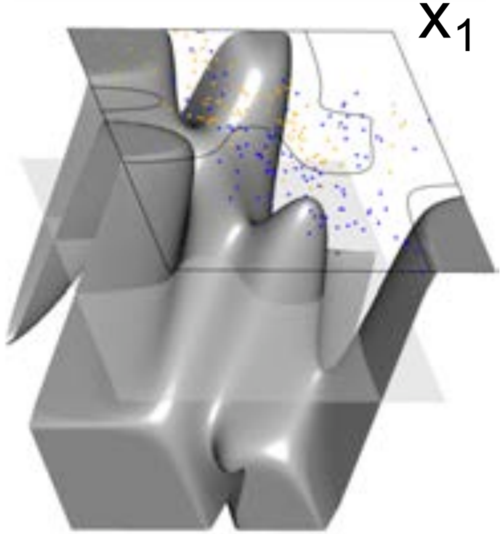
Milletari et al., V-Net: FCNNs for Volumetric Medical Image Segmentation, 3DV, 2016

Classification vs. Regression

Classification

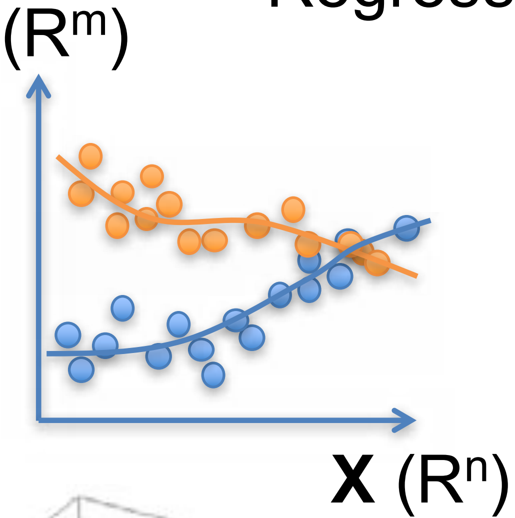
$$c = f(x)$$

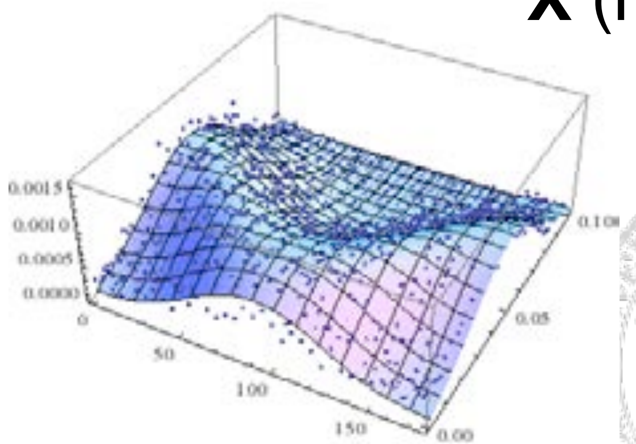




Regression

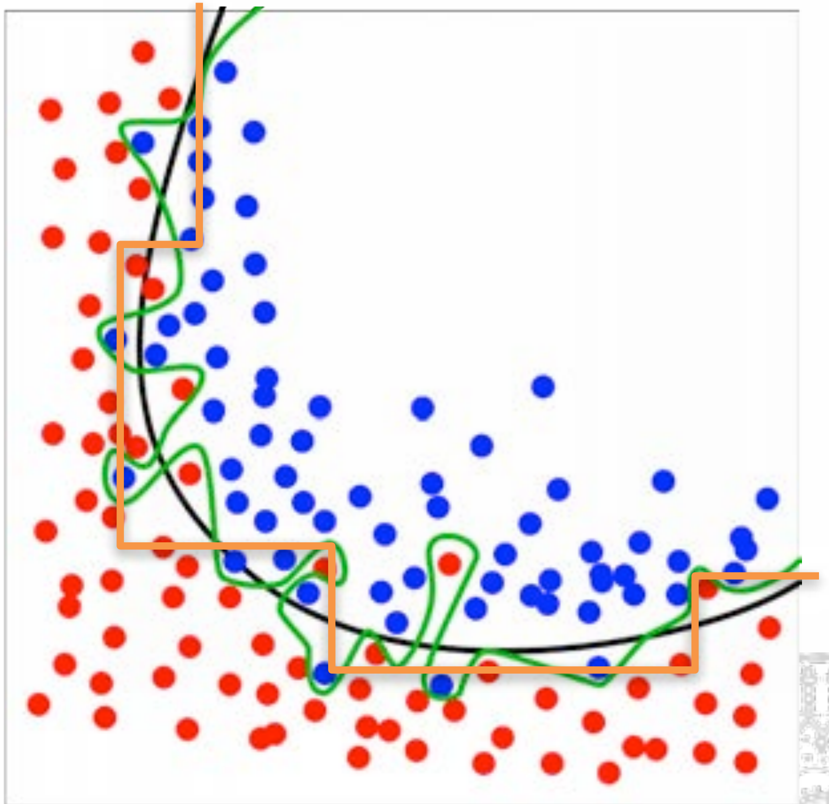
$$y = f(x)$$





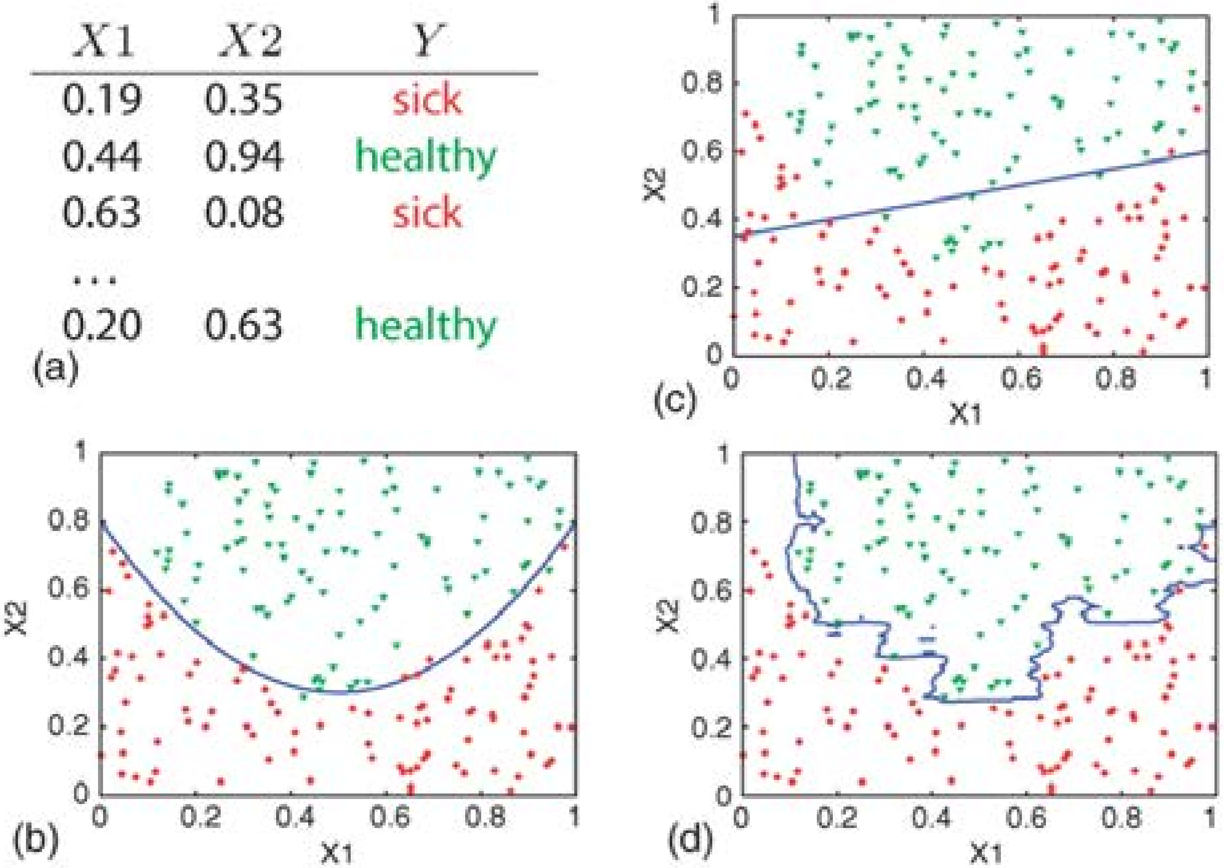
Challenges

- **Modeling the problem**
- Choosing the best learner
- Hyper-parameter tuning
- Optimization!!!
- Overfitting



https://upload.wikimedia.org/wikipedia/commons/1/19/Overfitting.svg

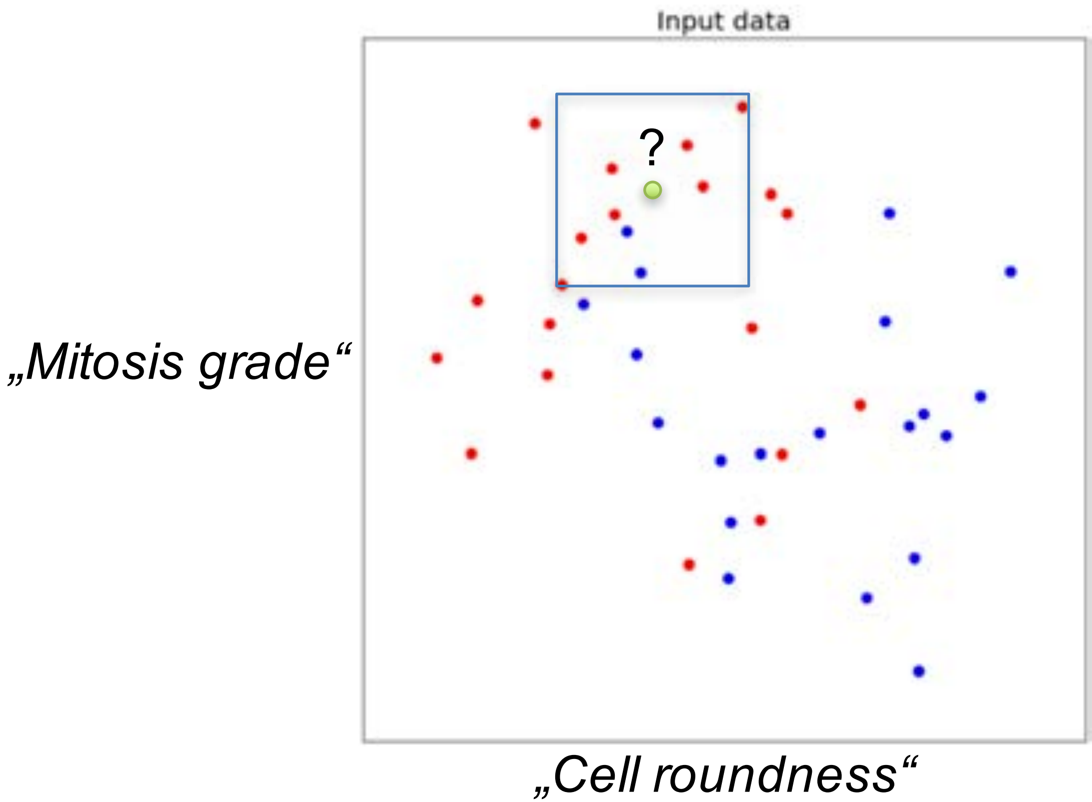
Overfitting

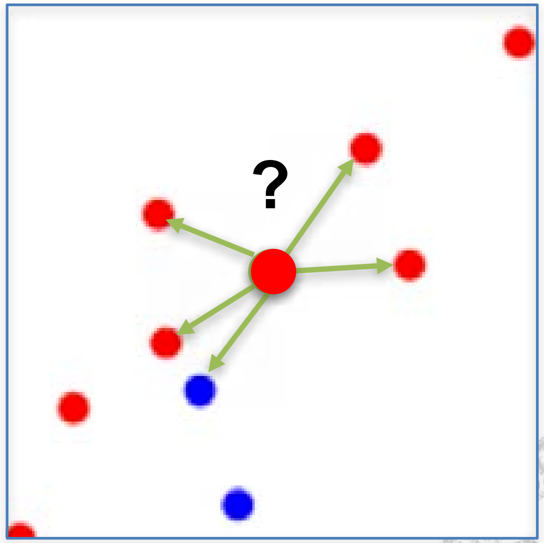


P. Geurts et al., Supervised learning with decision tree-based methods in computational and systems biology, Mol. BioSyst. 5 (2009) 1593–1605

K-Nearest Neighbors (kNN): "Intuitive idea"

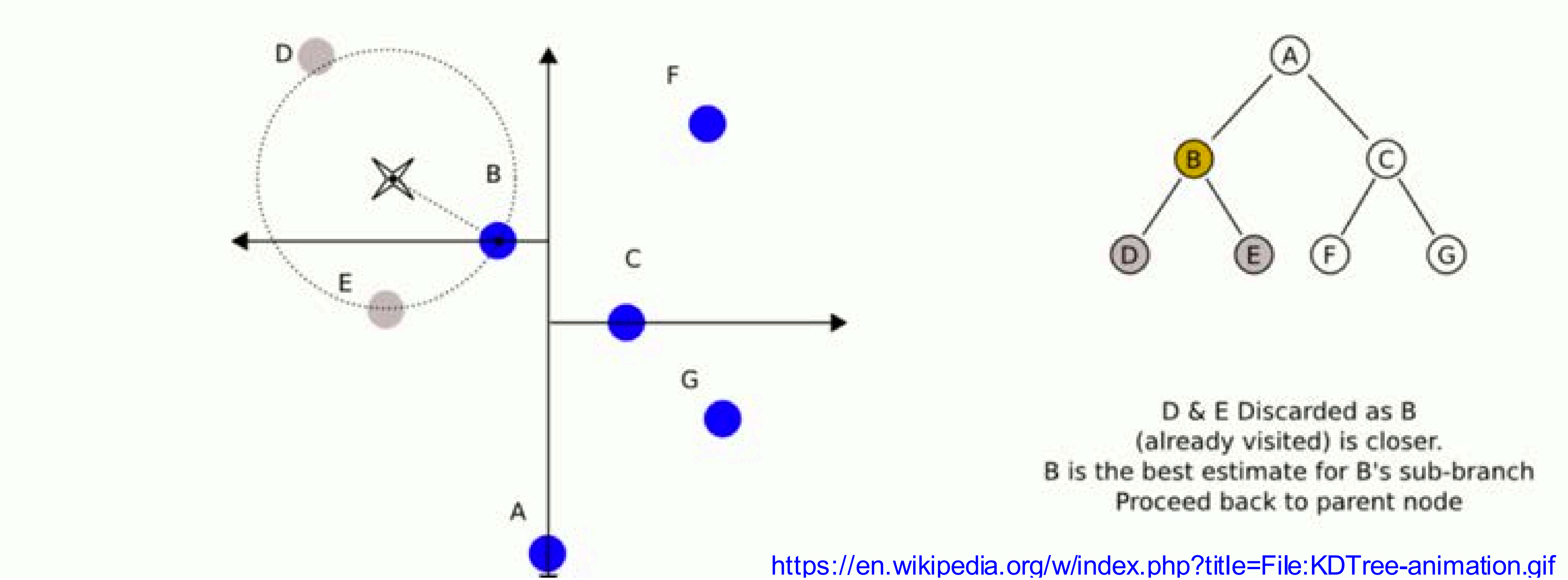
- *"A point's class should be determined by its nearest neighbors".*





K-Nearest Neighbors (kNN): Training

- kNN does not require an explicit “training”
- “Training” usually means partitioning the space to make NN searches faster and computationally more efficient

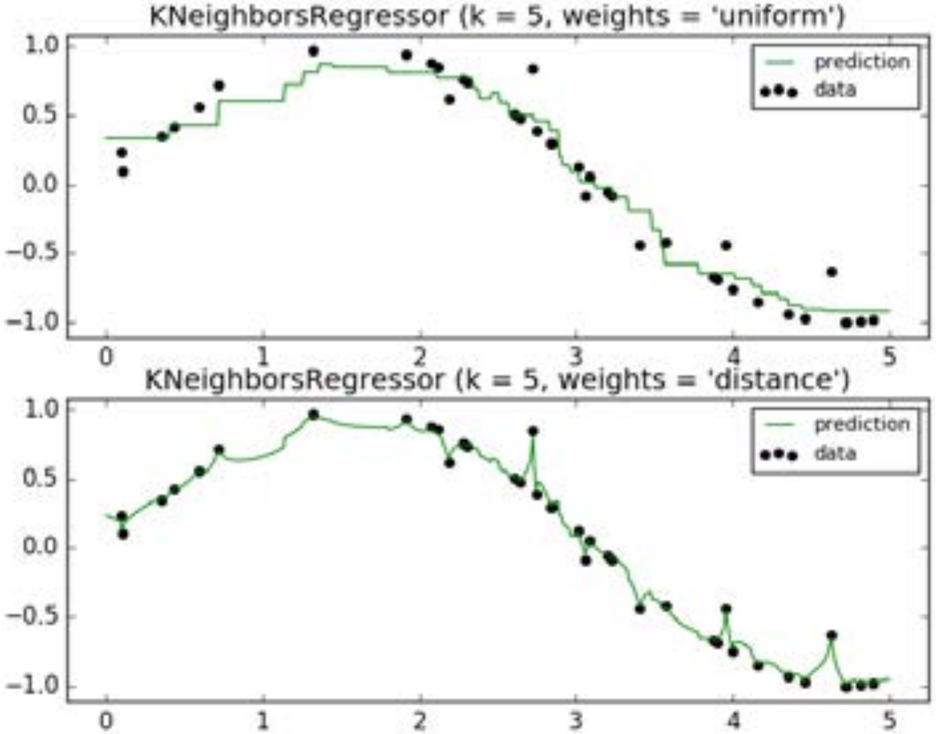


https://en.wikipedia.org/w/index.php?title=File:KDTree-animation.gif

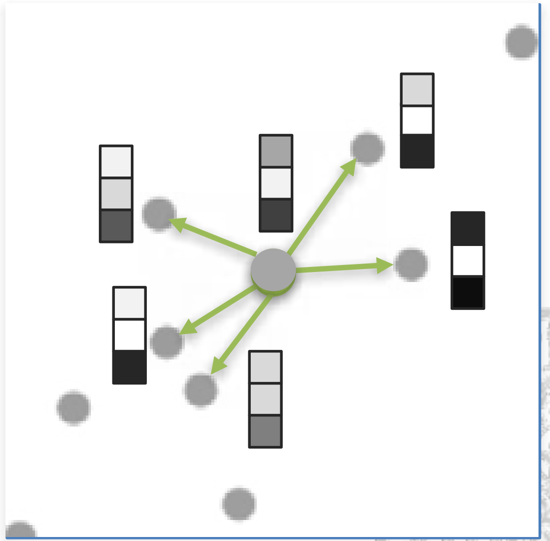
K-Nearest Neighbors (kNN): Testing (regression)

- Classification: majority vote of classes from neighbors
- Regression: average the targets from neighbors

1D → 1D



2D → 3D

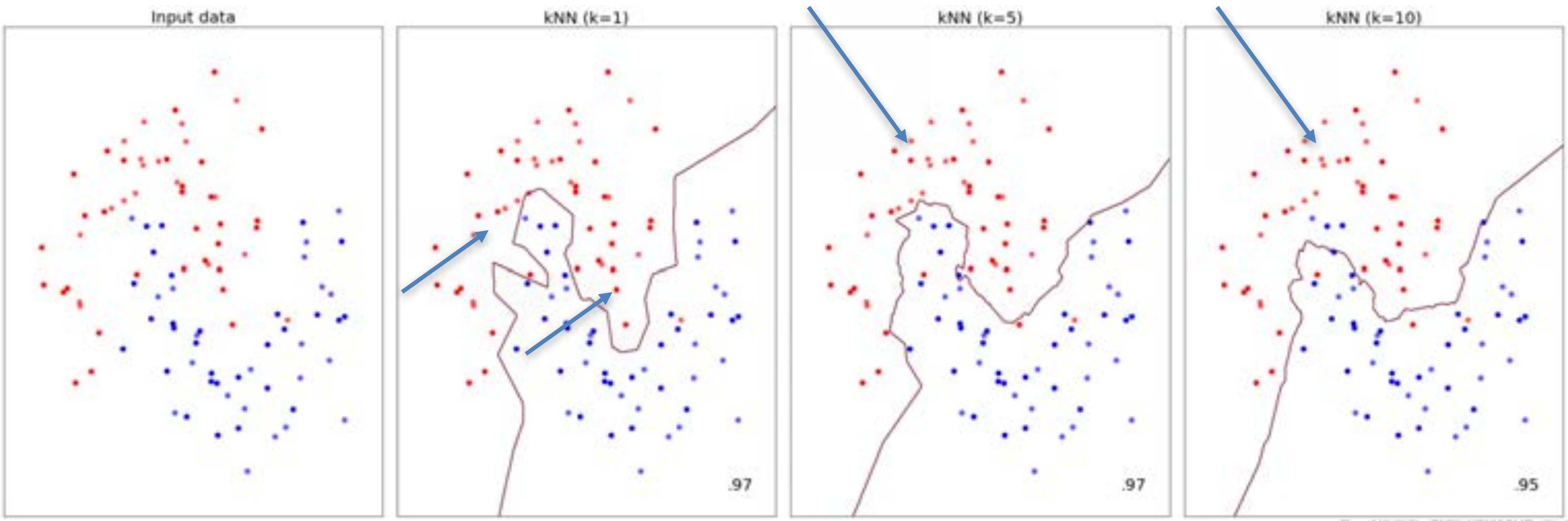


http://scikit-learn.org/stable/auto_examples/neighbors/plot_regression.html

K-Nearest Neighbors (kNN): Parametrization

- Classification (regression) depends on parameter *k*

k=1 k=5 k=10

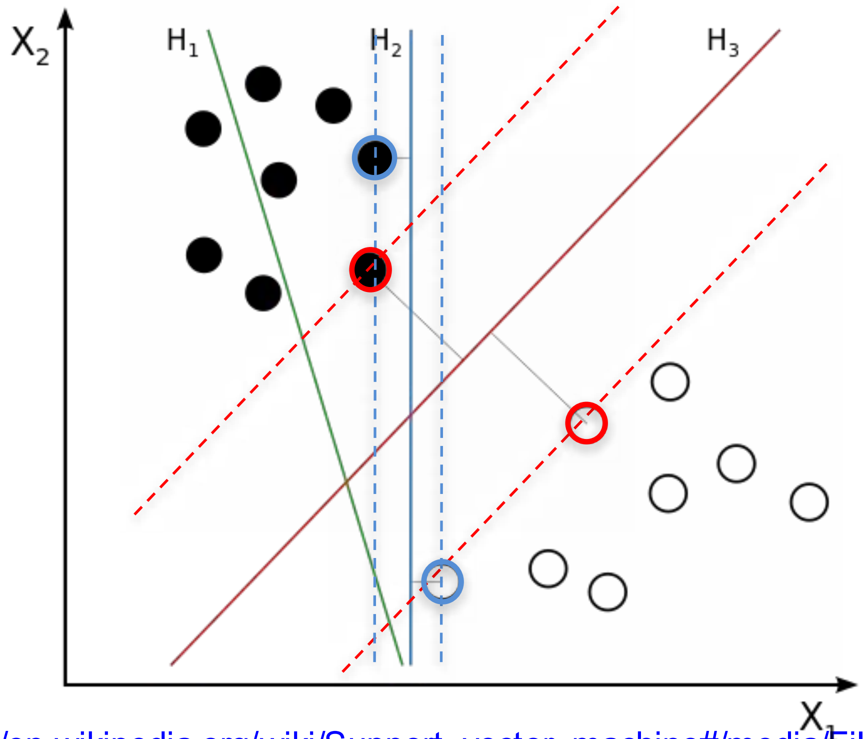


Support Vector Machines: "Intuitive idea"

- *"A few carefully selected key points at the boundary between two training classes should be enough to determine an unknown point's class."*

Careful selection → "optimally separating hyperplane"

Key points → "support vectors"



Classification at test time:

$$H(\bar{x}) \rightarrow \begin{array}{l} > 0 \ \rightarrow class\ 1 \\ < 0 \ \rightarrow class\ 0 \end{array}$$

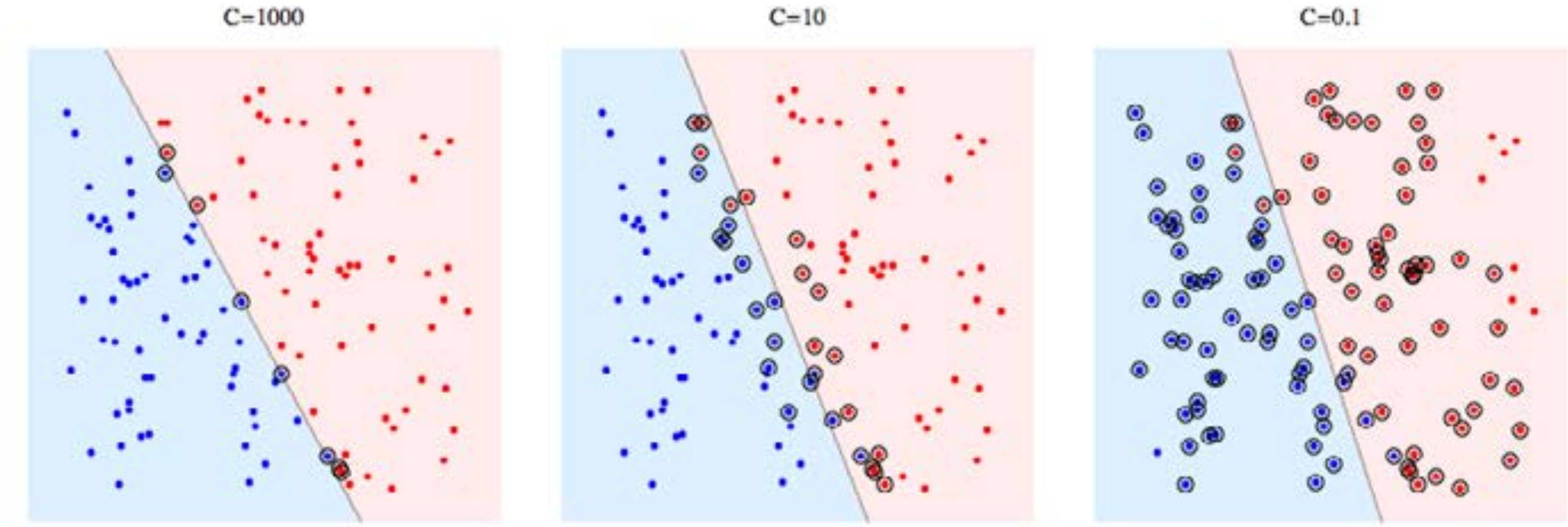
Finding parameters for H is based on mathematical constraints and involves an iterative optimization given all training data to find support vectors.

Image: https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_separating_hyperplanes_(SVG).svg

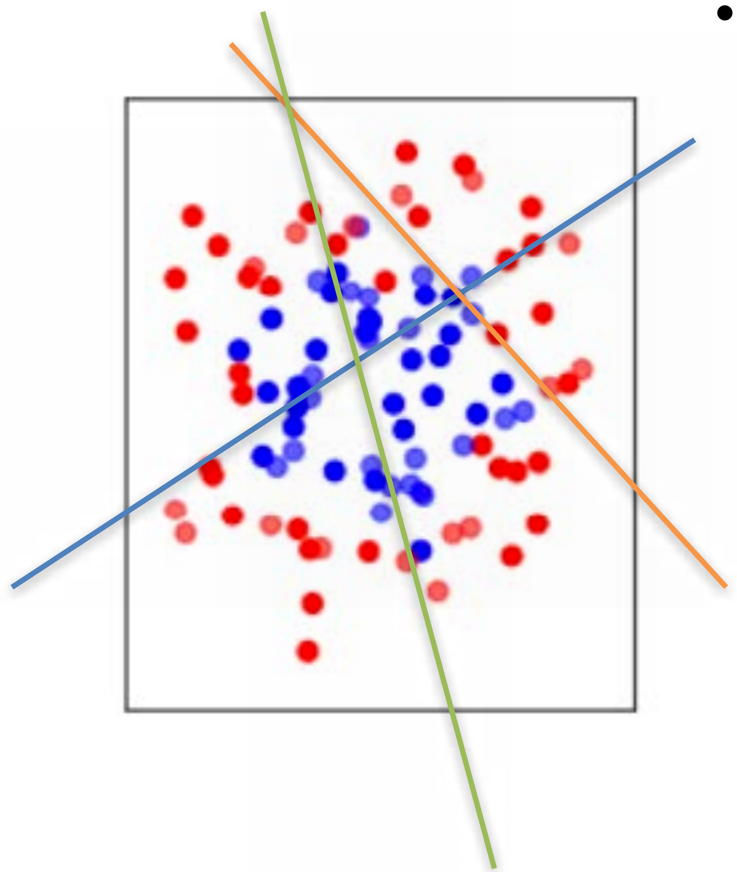
For a really beautiful derivation of SVM, watch MIT AI course 6.034 (2010) on https://youtu.be/_PwhiWxHK8o

Support Vector Machines: Training with soft-margin

- **Soft-margin:** Given non-separability, how do we still find the separating plane?
- C: complexity parameter which determines how much "slack" is given to support vectors (especially mis-classified ones)
- C large: find a hard separation plane
- C small: find a soft separation plane, while allowing many mis-classifications (useful when we do not "trust" the distribution of our training data)

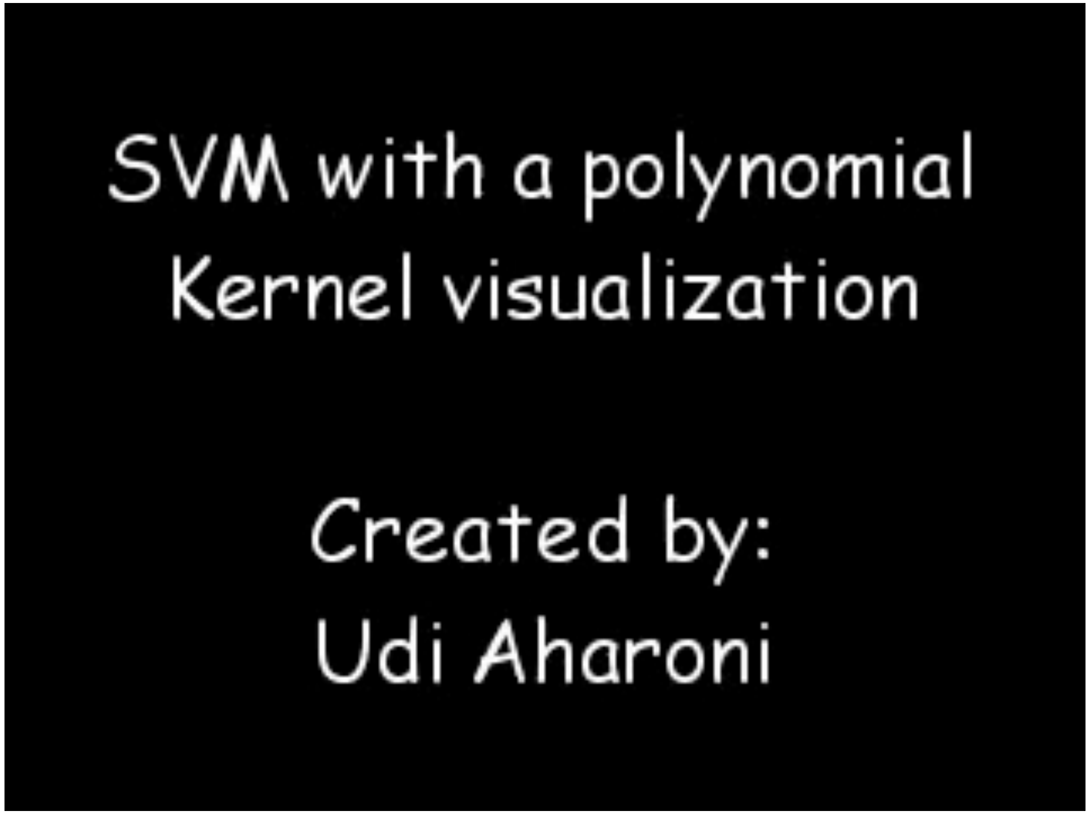


Support Vector Machine: Non-linear classification?



- No linear hyperplane (i.e. no straight line in 2D) will be able to separate these two classes…

Support Vector Machine: Non-linear classification with the “Kernel trick”

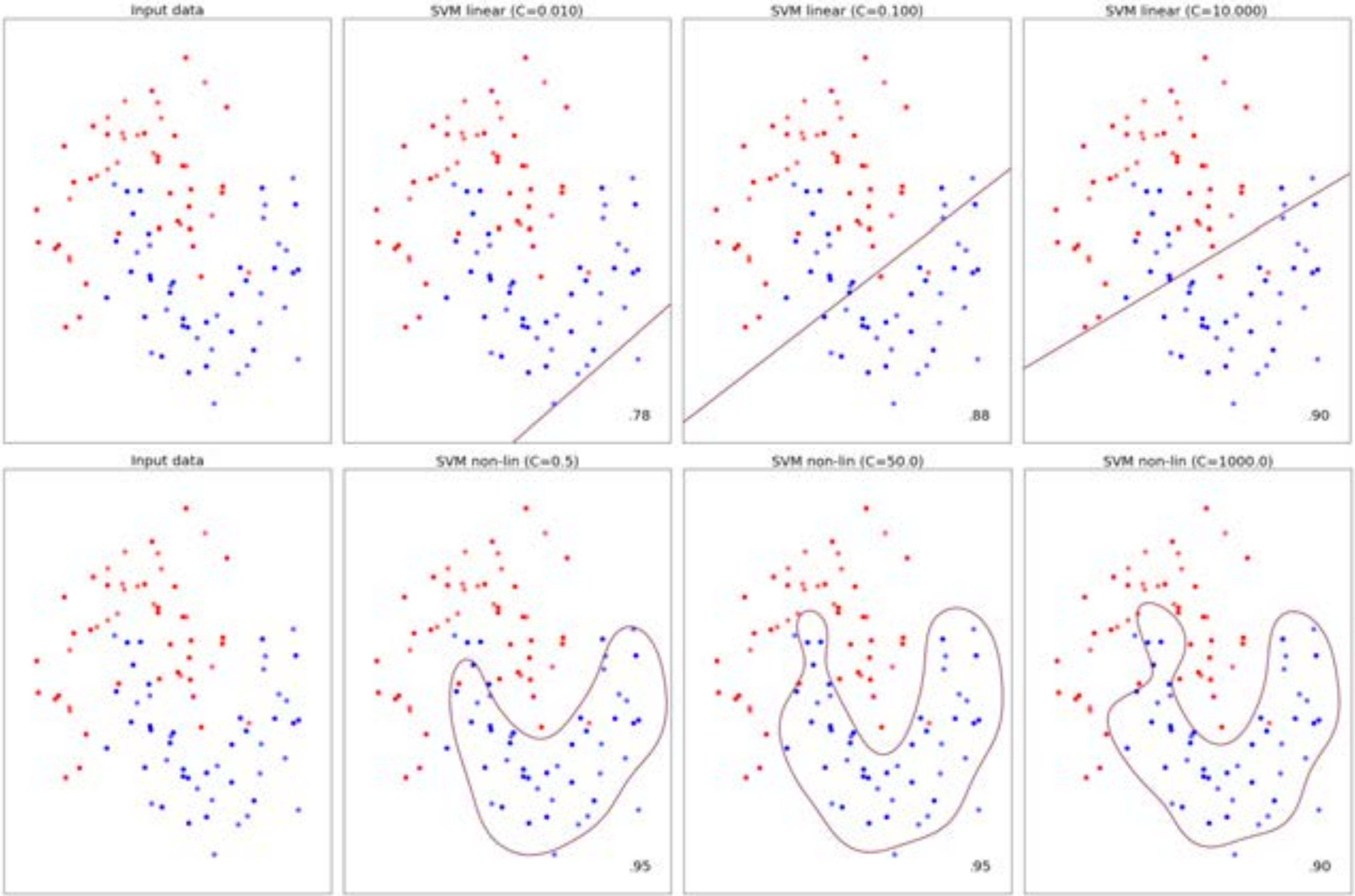


https://www.youtube.com/watch?v=3liCbRZPrZA

Support Vector Machines: Parametrization

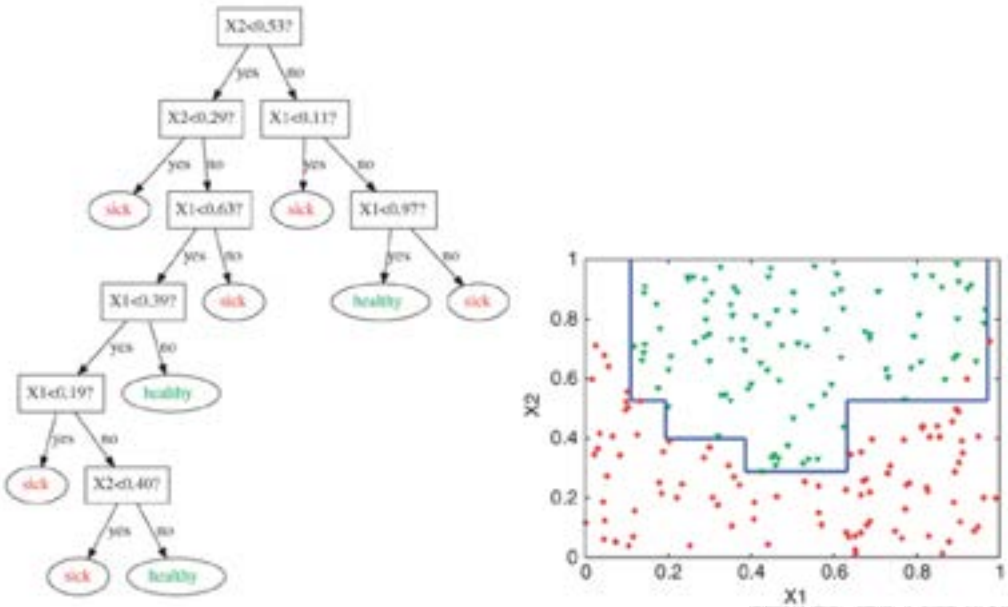
Linear SVM

Non-linear SVM
(Gaussian kernel)



Decision trees and Random Forests: Principle

- Intuitive idea behind decision trees:
  "The class of an unknown point can be determined by a series of rules/decisions"
- Problem: Single decision trees overfit quickly (i.e. generalize badly)
- Intuitive idea behind Random Forests:
  "Improve classification by training several decision trees on different aspects of our training data, and combine them into one common ensemble – a 'forest'."
- ***"Where one tree fails, the others do well."***



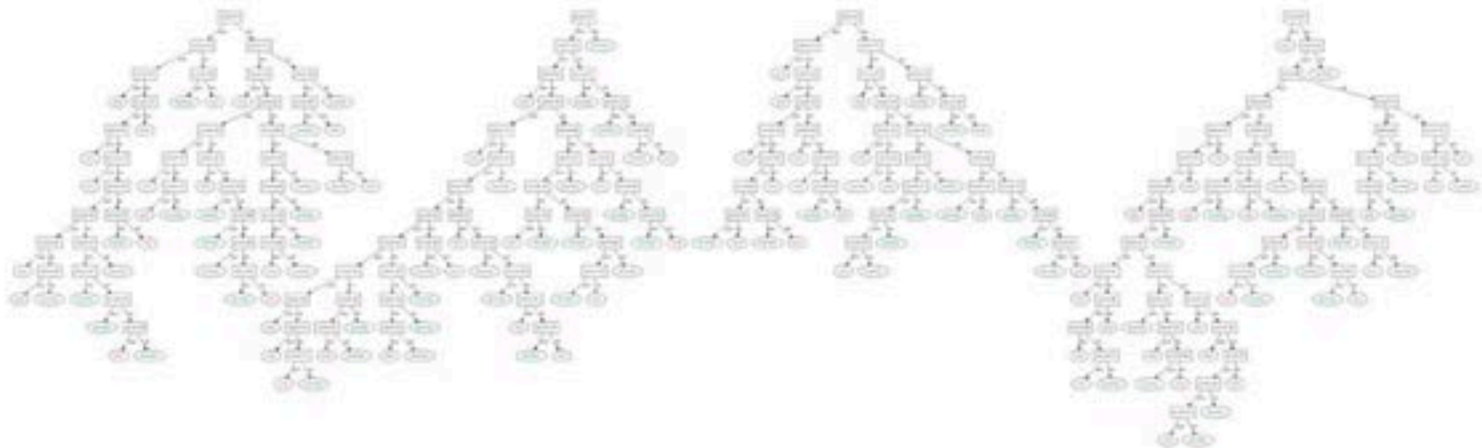
P. Geurts et al., Supervised learning with decision tree-based methods in computational and systems biology, Mol. BioSyst. 5 (2009) 1593–1605

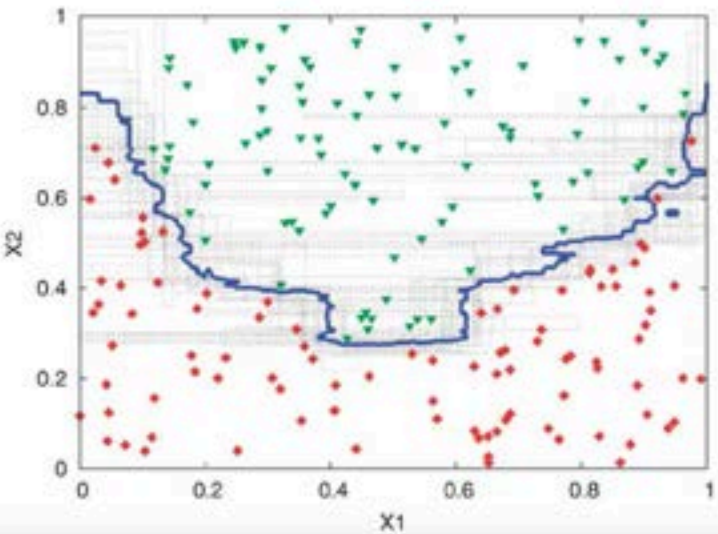
Random Decision Forests: Principle

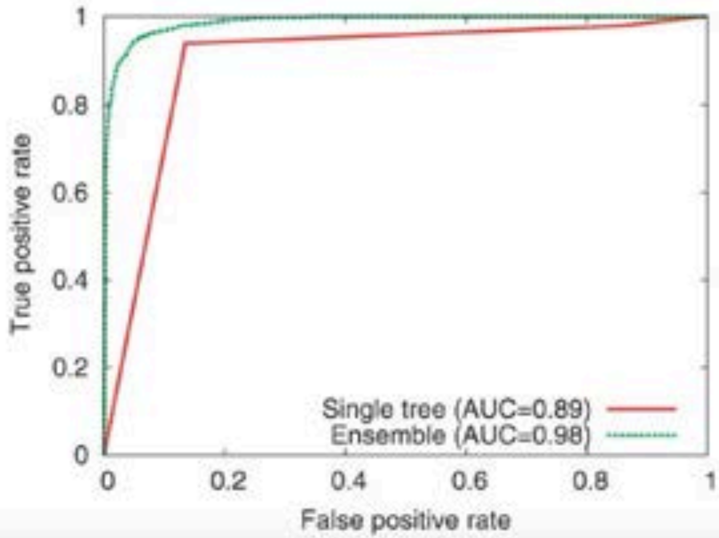
Randomizable training aspects:

- Training data
- Subset of dimensions at each split
- Random split decisions (extremely randomized trees)

Randomization provides excellent regularization!







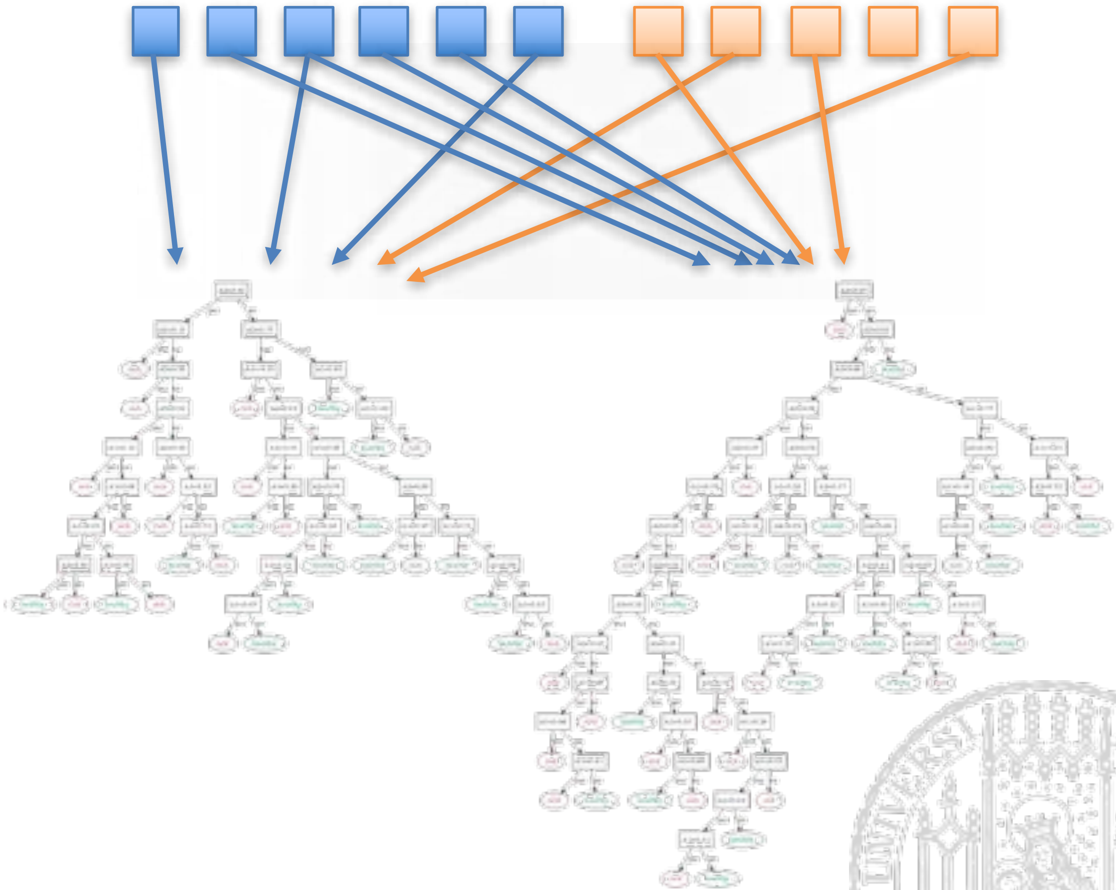
P. Geurts et al., Supervised learning with decision tree-based methods in computational and systems biology, Mol. BioSyst. 5 (2009) 1593–1605

Random Decision Forests

Randomizable training aspects:

- Training data
- Subset of dimensions at each split
- Random split decisions (extremely randomized trees)

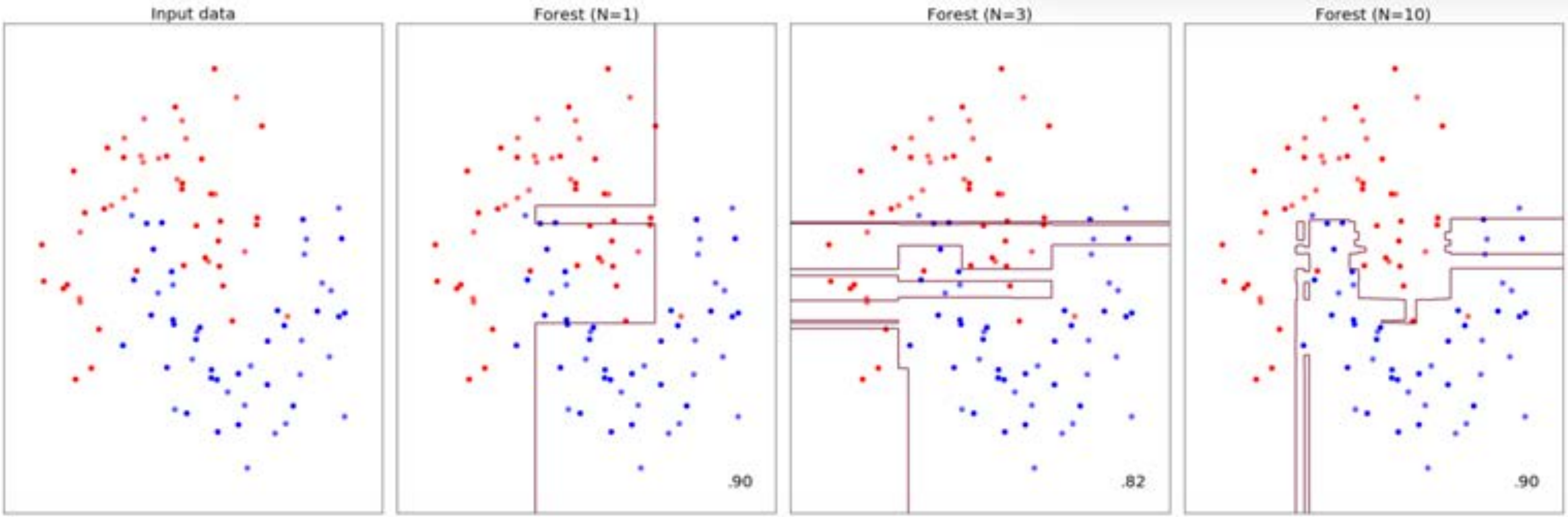
Randomization is a very effective means to achieve regularization!



Random Decision Forests: Parametrization

- Classification depends e.g. on number of trees N

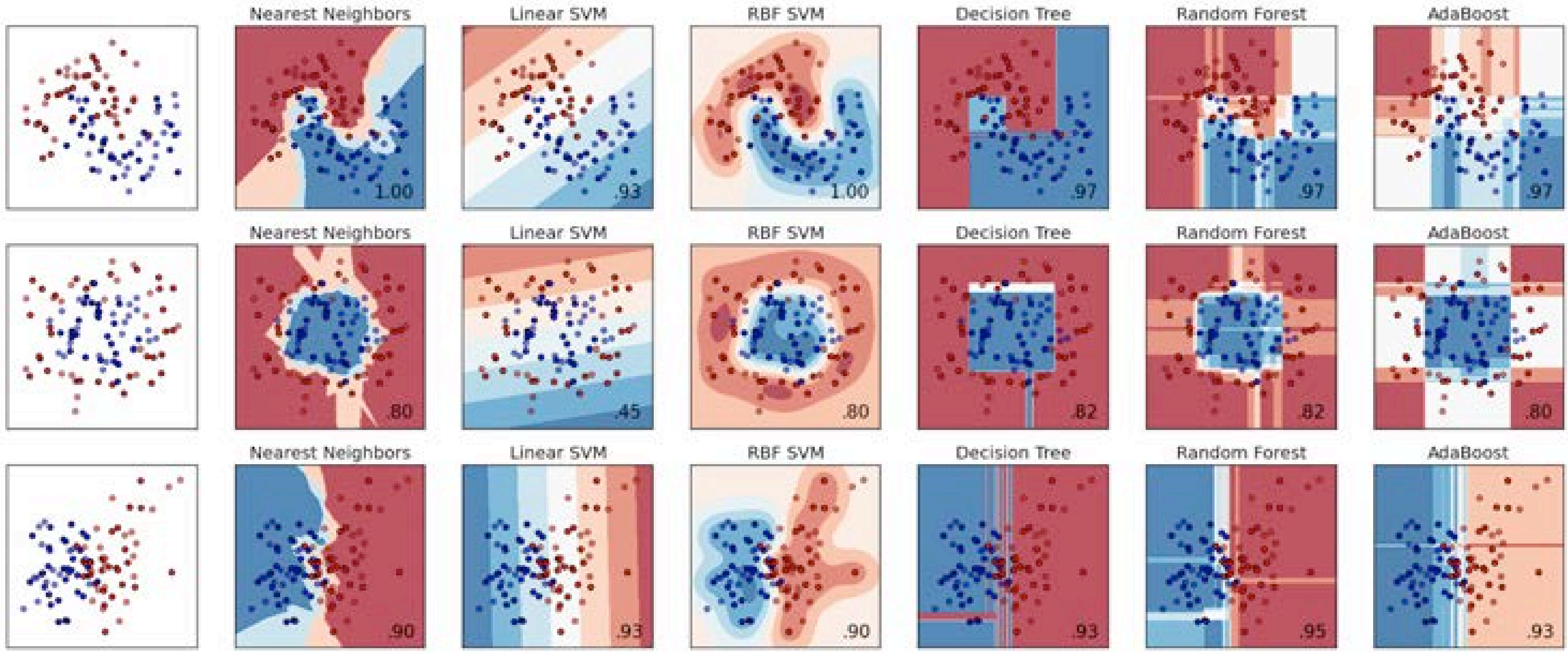
N=1 N=3 N=10



Which classifier/regressor to choose?

	kNN	SVM	Random Forest
Required dimensionality	Small	Medium	Very high
Required dataset size	Medium (acceleration due to parcellation e.g. with kdtrees)	Small (training on large datasets can be painfully slow)	Very large
Efficiency		Very slow to train on >lab-size datasets	Highly efficient, parallelizable

Many more choices…



Adapted from: http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

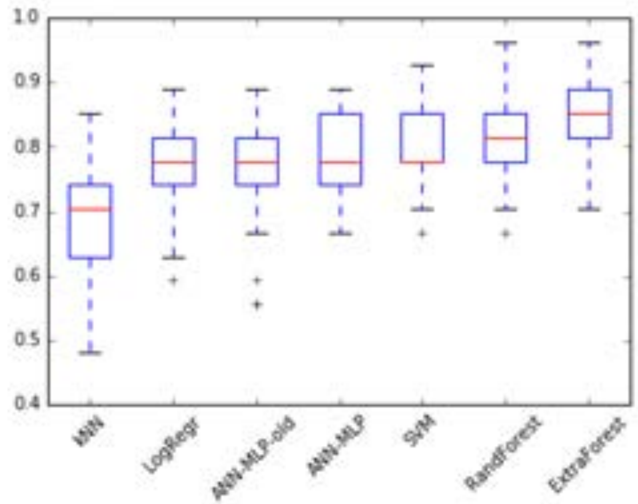
KLINIKUM

DER UNIVERSITÄT MÜNCHEN

DSGZ

Statistical robustness via cross-validation

- Achieve robustness in machine learning validation via cross-validation or
- Splitting into training-/validation-/test-dataset
- Give confidence intervals for classification accuracy
- Give a p-value via label permutation tests



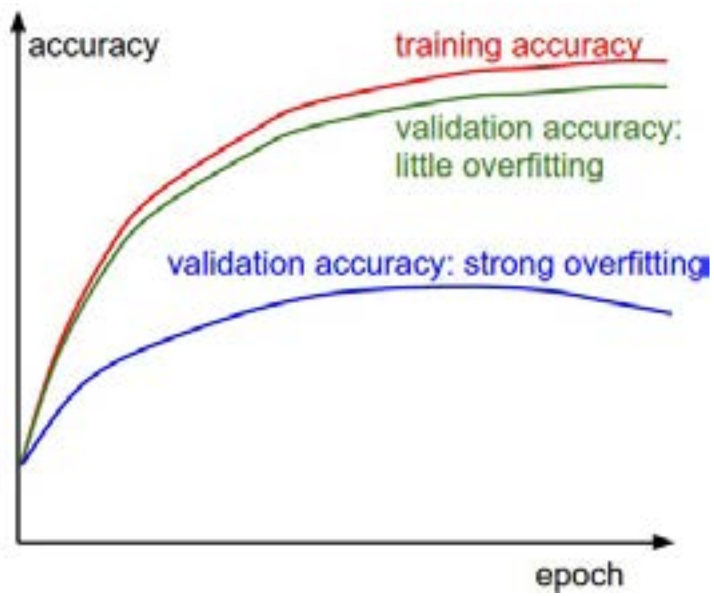
	Dataset splitting							
Fold 1								
Fold 2								
Fold 3								
Fold 4								

Train	Test

Statistical robustness via validation and test set

- Useful when training of folds takes very long
  - Complex models, e.g. deep neural networks
  - Extremely large datasets, e.g. ImageNet in Computer Vision

	Dataset splitting		
Single fold	Training set	Validation set	Test set

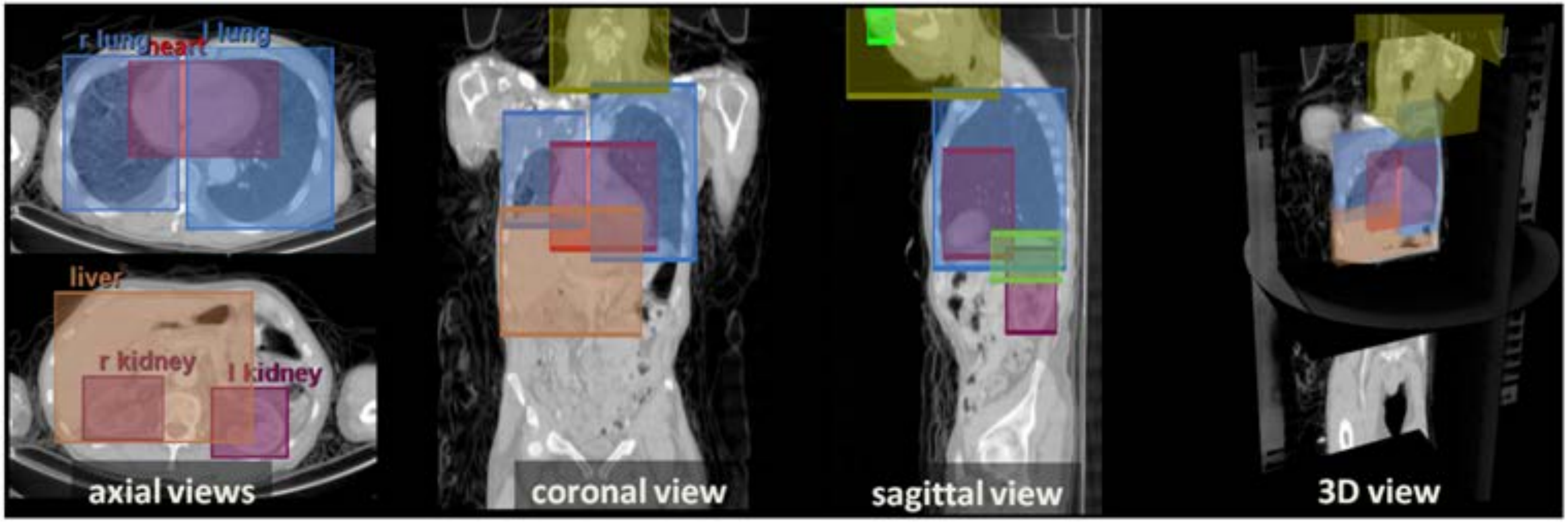


The validation set is used

- to observe generalizability of the model to unknown data
- to tune the parameters of the network
- to observe convergence behaviour of the optimization

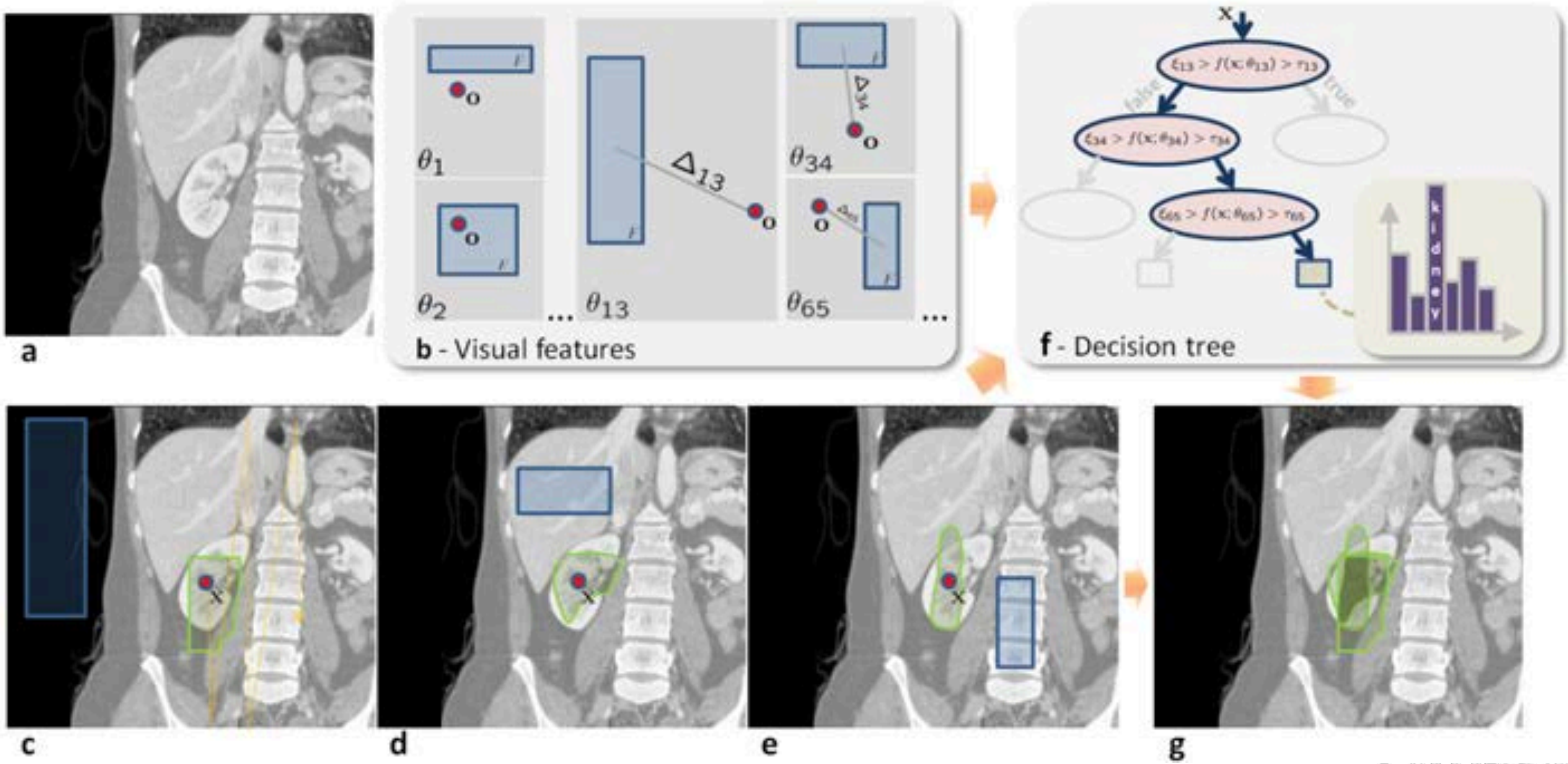
Image from: CS321n
http://cs231n.github.io/neural-networks-3/

Machine learning in medical imaging: Anatomy localization with random forest classification



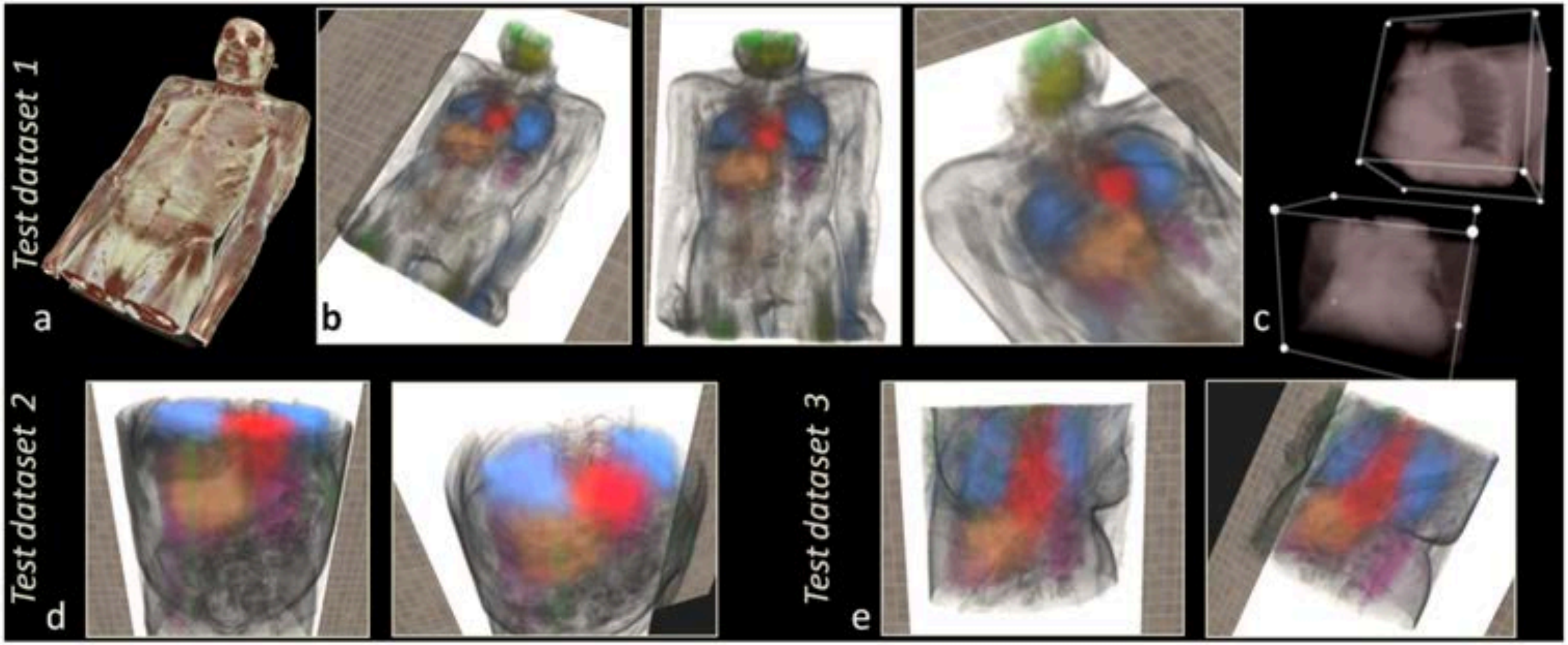
Criminisi et al., Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes, MICCAI Workshop on Probabilistic Models for Medical Image Analysis (MICCAI-PMMIA), Sept. 2009

Machine learning in medical imaging: Anatomy localization with random forest classification



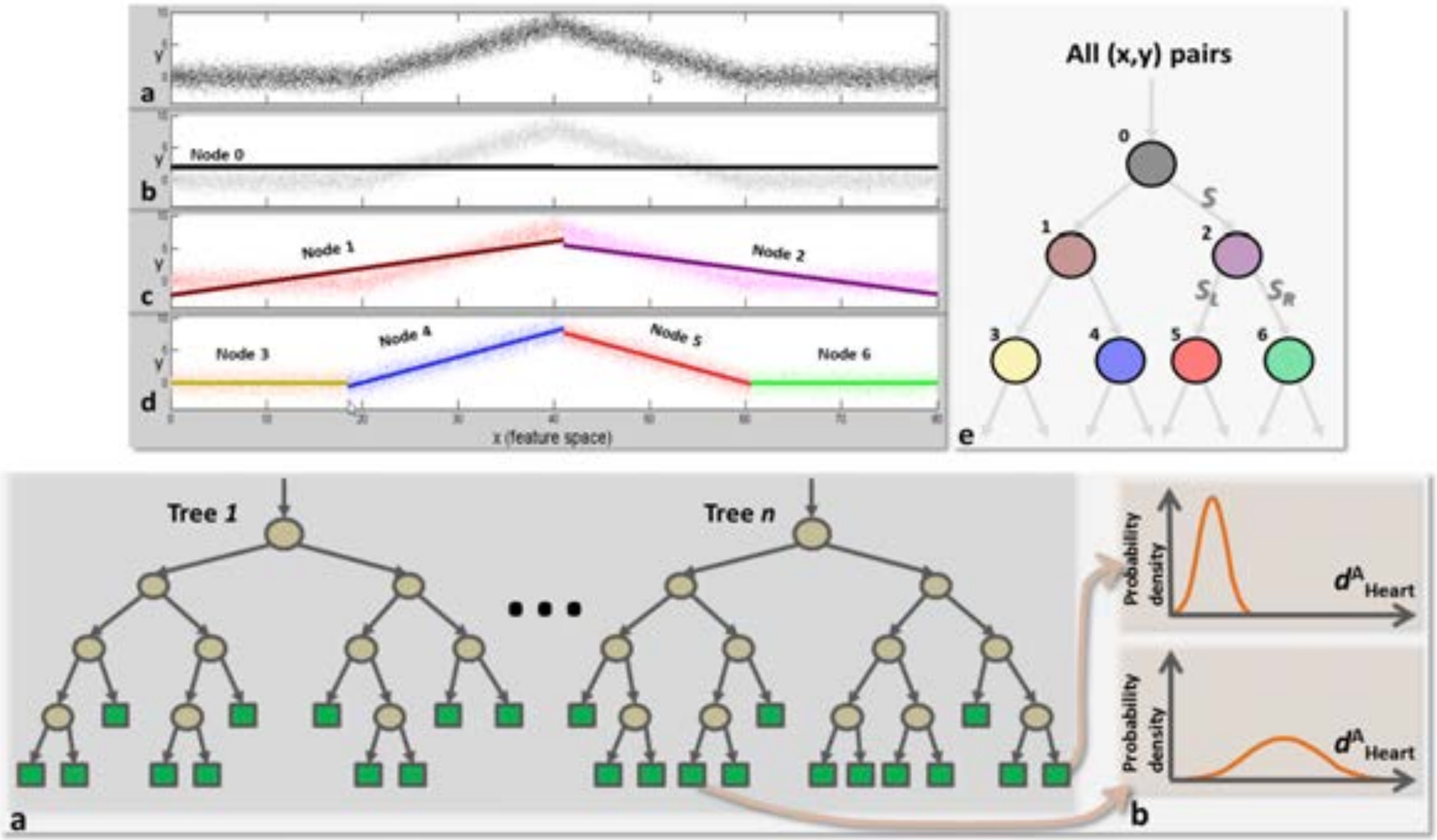
Criminisi et al., Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes, MICCAI Workshop on Probabilistic Models for Medical Image Analysis (MICCAI-PMMIA), Sept. 2009

Machine learning in medical imaging: Anatomy localization with random forest classification



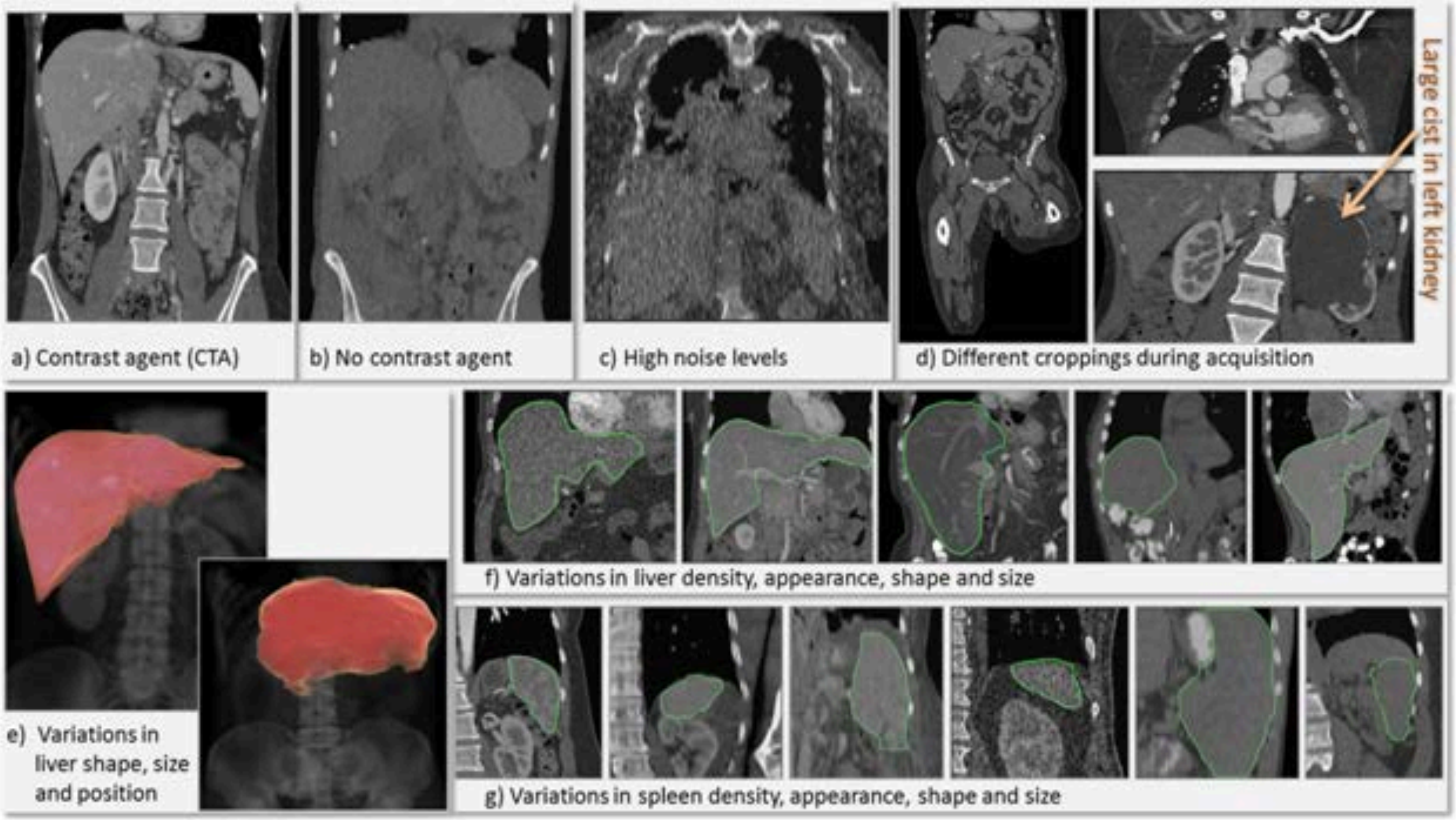
Criminisi et al., Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes, MICCAI Workshop on Probabilistic Models for Medical Image Analysis (MICCAI-PMMIA), Sept. 2009

Machine learning in medical imaging:
Anatomy localization with random forest regression



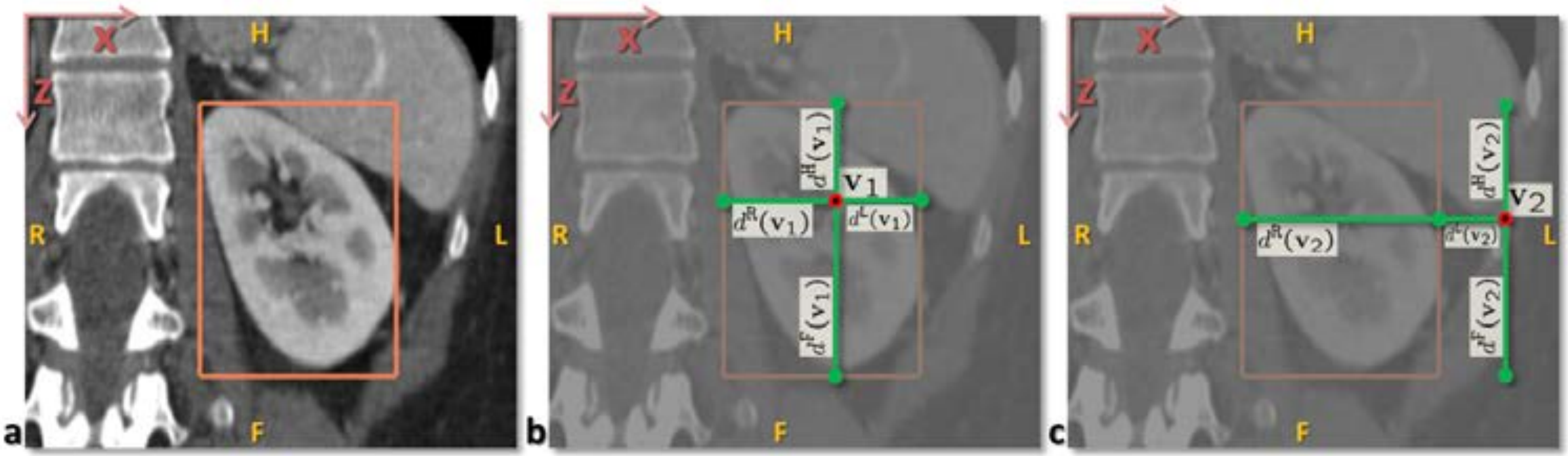
Criminisi et al., Regression Forests for Efficient Anatomy Detection and Localization in CT Studies, International MICCAI Workshop on Medical Computer Vision (MICCAI-MCV), LNCS vol. 6533, pp. 106-117, 2010

Machine learning in medical imaging: Anatomy localization with random forest regression



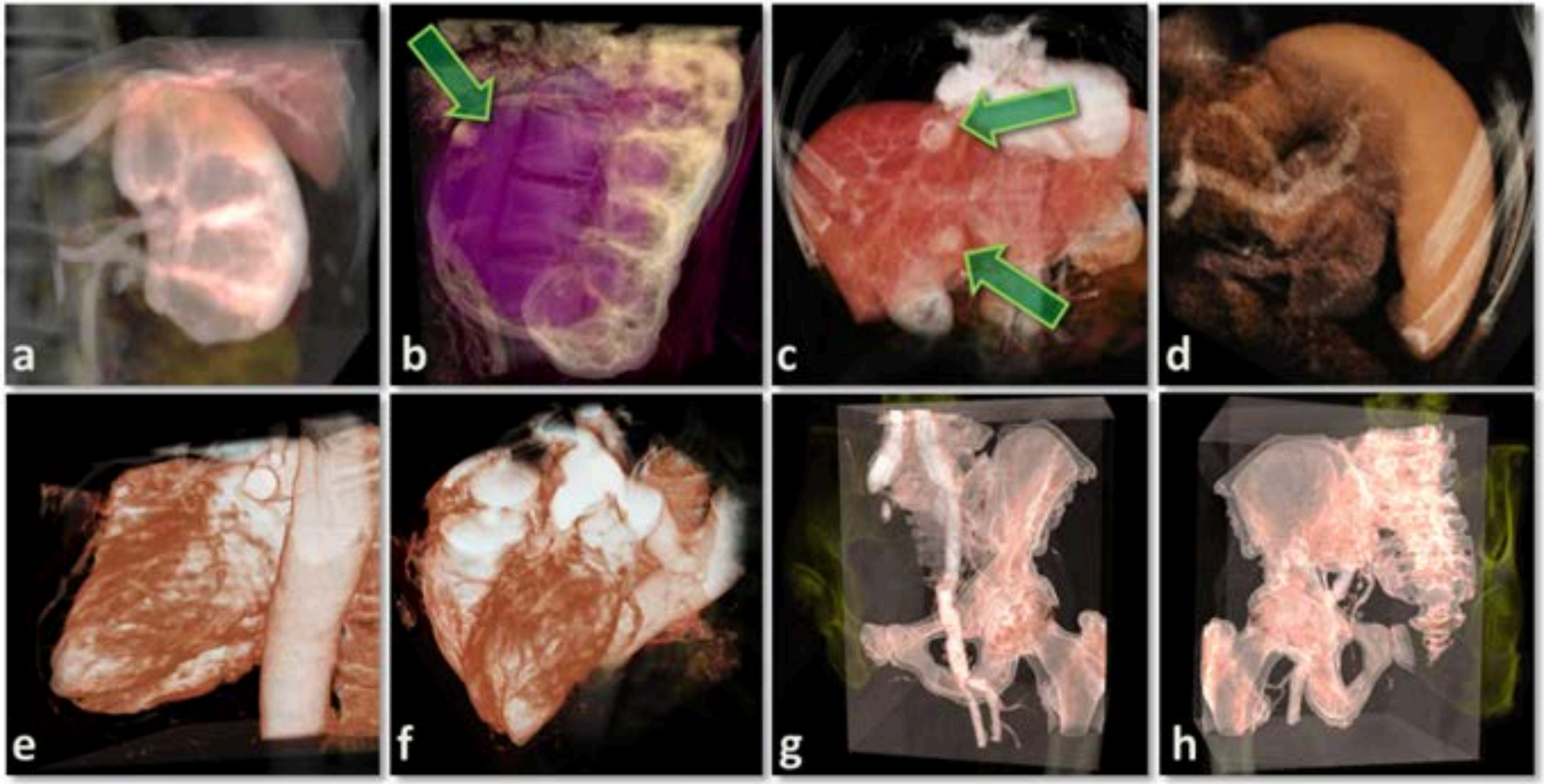
Criminisi et al., Regression Forests for Efficient Anatomy Detection and Localization in CT Studies, International MICCAI Workshop on Medical Computer Vision (MICCAI-MCV), LNCS vol. 6533, pp. 106-117, 2010

Machine learning in medical imaging: Anatomy localization with random forest regression



Criminisi et al., Regression Forests for Efficient Anatomy Detection and Localization in CT Studies, International MICCAI Workshop on Medical Computer Vision (MICCAI-MCV), LNCS vol. 6533, pp. 106-117, 2010

Machine learning in medical imaging: Anatomy localization with random forest regression



Criminisi et al., Regression Forests for Efficient Anatomy Detection and Localization in CT Studies, International MICCAI Workshop on Medical Computer Vision (MICCAI-MCV), LNCS vol. 6533, pp. 106-117, 2010

DSGZ