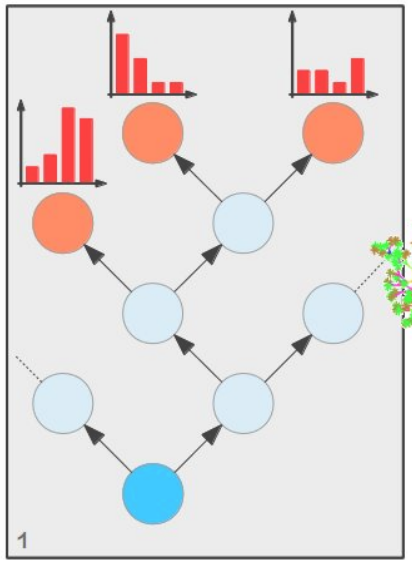
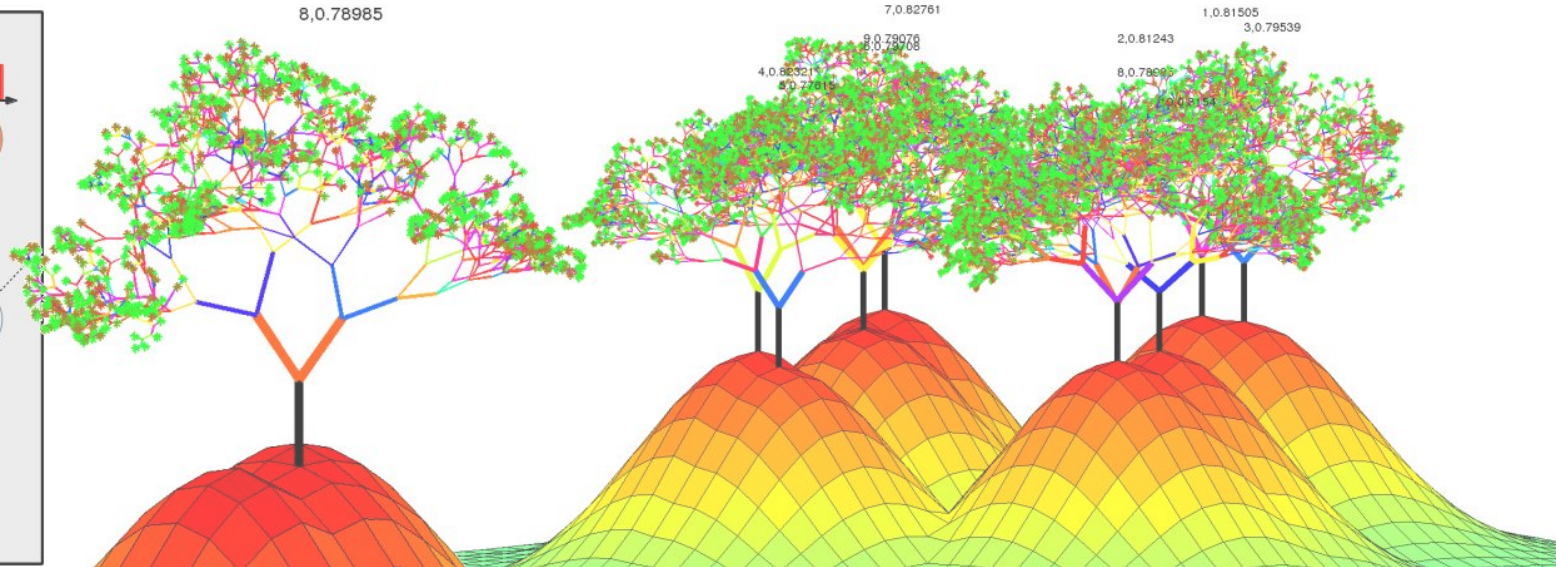
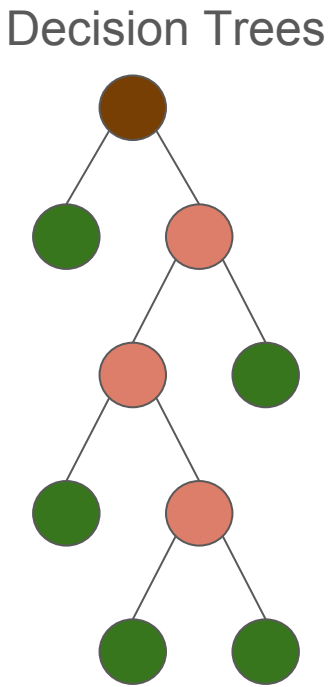
Random Forest

Richard Olney, Nick Knowles
{olneyr2, knowlen}@wwu.edu



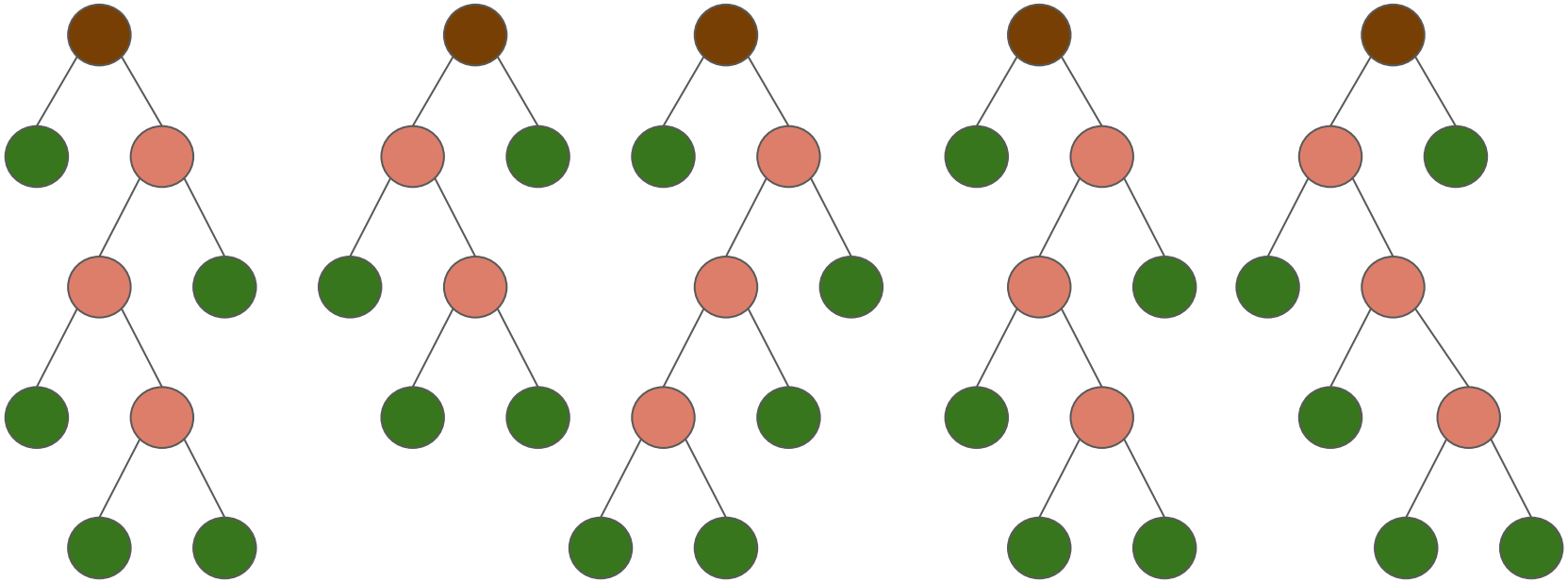


Ensemble Model



Ensemble Model

Decision Forest



Tree bagging

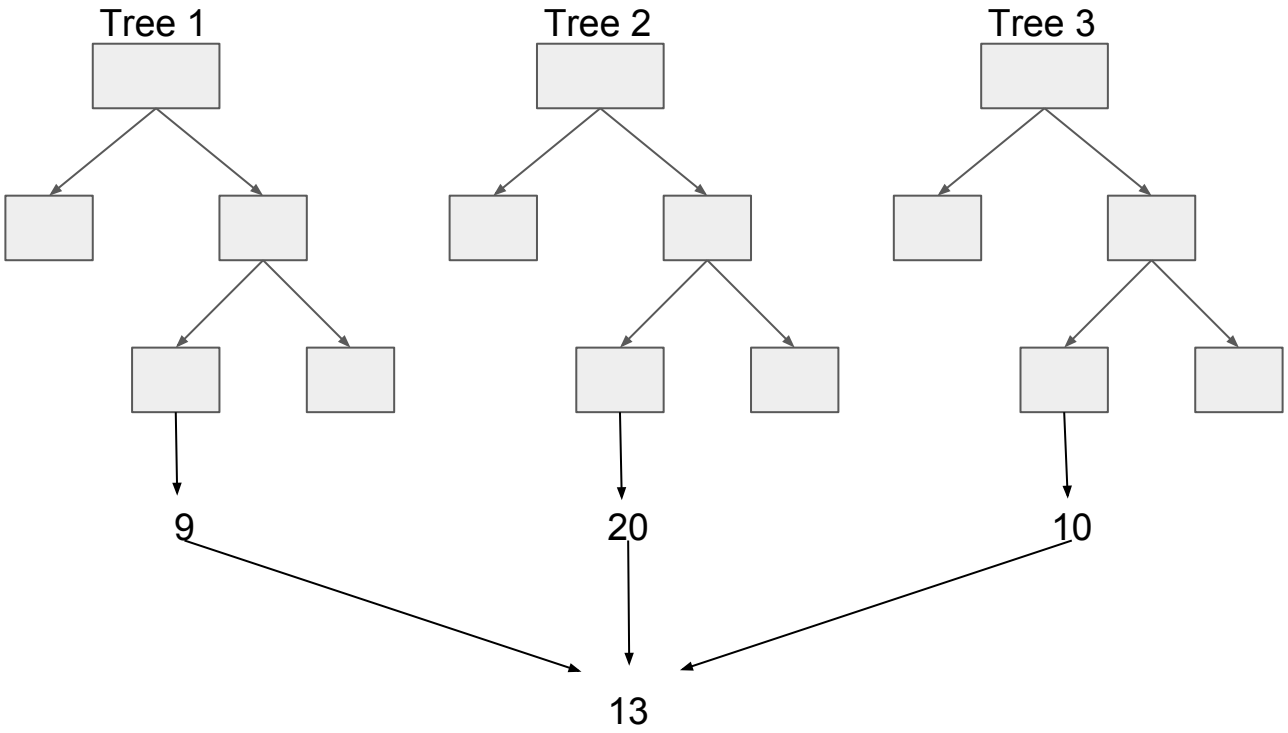
Training

- Random forest is a supervised learning algorithm.
- From a data set of size N, select with replacement N training points for each tree in the forest.
- Grow/Train each tree on a subset of m randomly selected features.
- As the number of trees in the forest grows the more accurate it becomes.

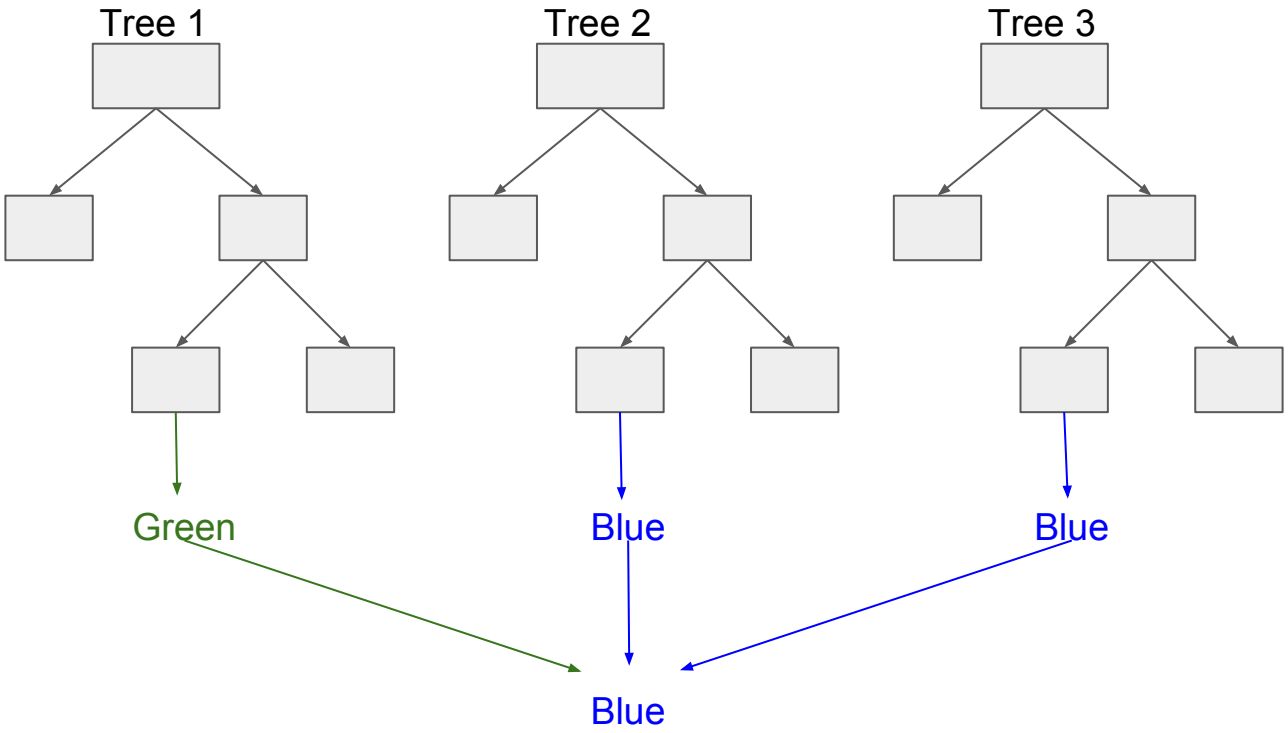
Prediction

- For evaluation run the test sample through each decision tree in the forest and assess the results.
  - For a **regression** forest the prediction is the average of the predictions of all the trees.
  - For a **classification** forest the prediction is the majority class produced by the forest.

Regression Forest



Classification Forest



Feature bagging

Randomly select a subset of m features to train each tree on.

Then use your favorite decision tree algorithm.

Pseudo Code

- While less than the number of trees you want
  - Sample a new training set with replacement
  - Train a new tree using feature bagging
- End

Why RF? Key Properties

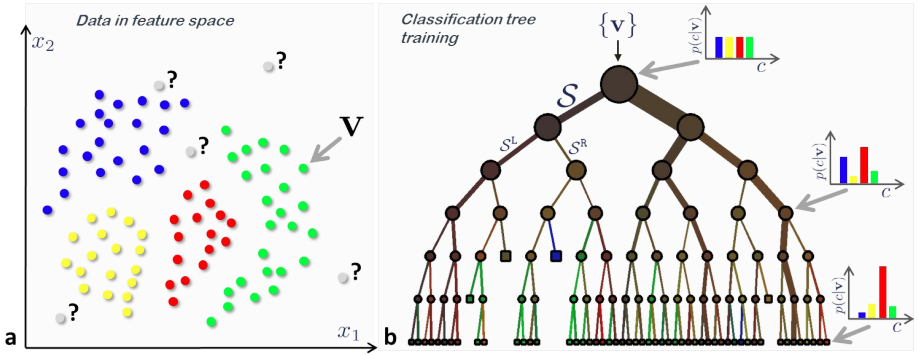
Low variance (hard to overfit)

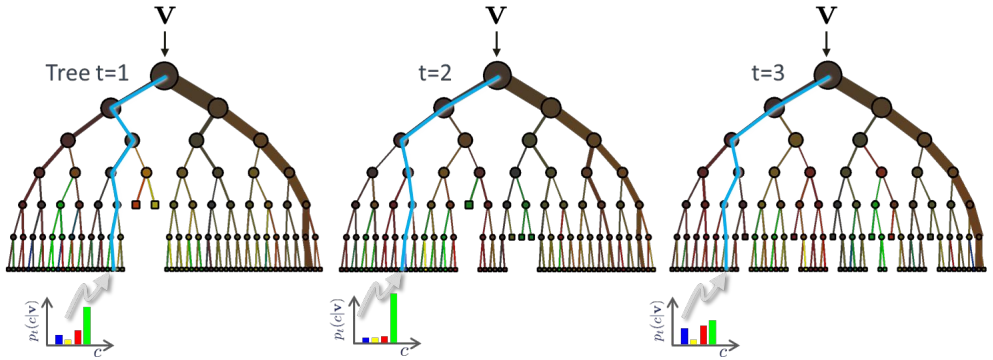
Robust to noise

Can handle large, high dimensional data

Trivial to parallelize

Can explain what features are most important towards some objective

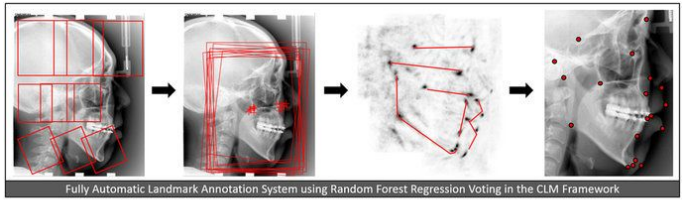


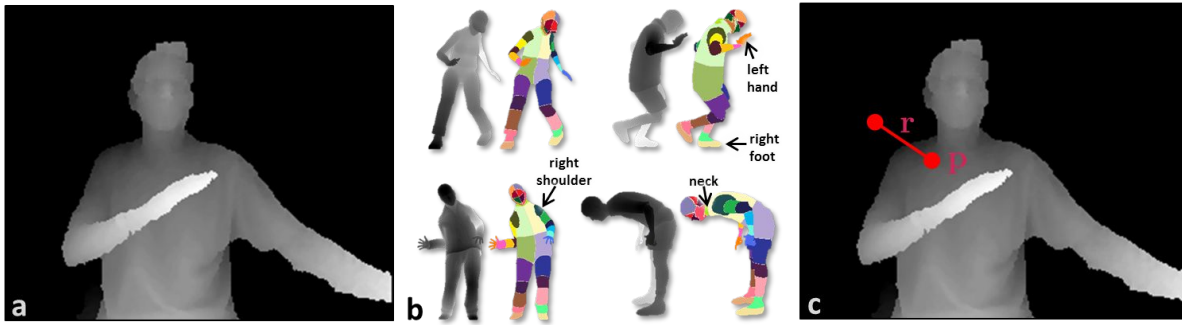


Applications In The World

Classification, Regression, Density Estimation, Clustering, Manifold Learning, ...

Quantitative finance
Bioinformatics
Natural language processing
Recommender Systems
Image, audio recognition
cont...





[Shotton et al 2013]

Hyperparameters

Forest Size: number of trees to bag.

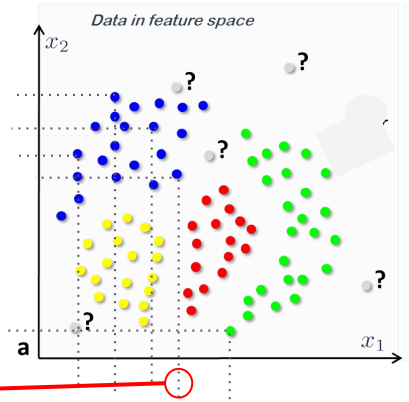
Max Features: Maximum number of features to consider when splitting a node.

Tree Depth: Maximum allowed distance from root node to leaf.

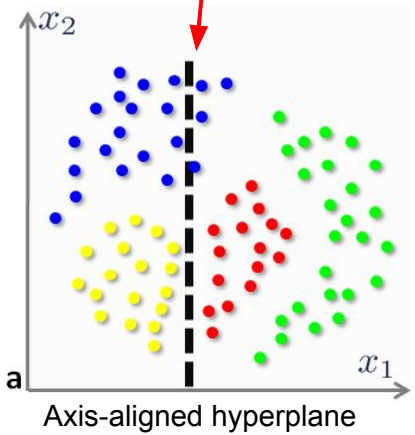
Weak Learner: Decision model (geometric primitive) that carves up input space.

Hyperparameters

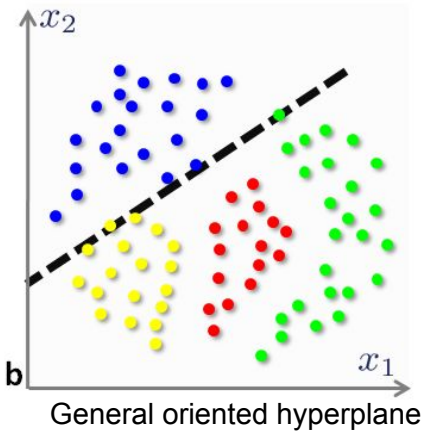
Weak Learner



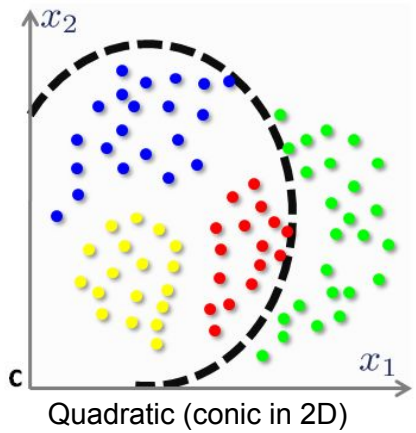
Selected decision boundary



Axis-aligned hyperplane



General oriented hyperplane



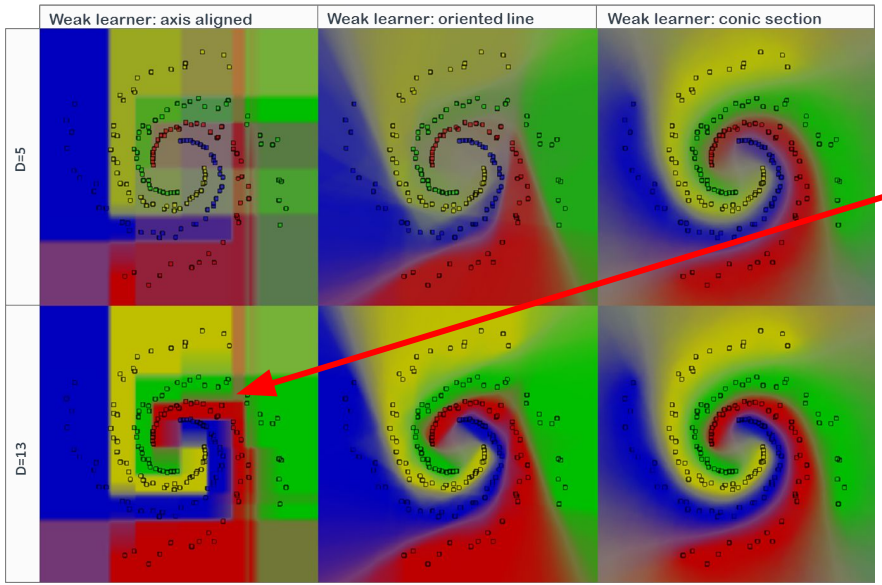
Quadratic (conic in 2D)

[Criminisi et al 2011]

Hyperparameters

Weak Learner

More complicated models produce higher quality of output confidence, but are less efficient to compute.



Blocky regions tend to indicate bad generalization

[Criminisi et al 2011]

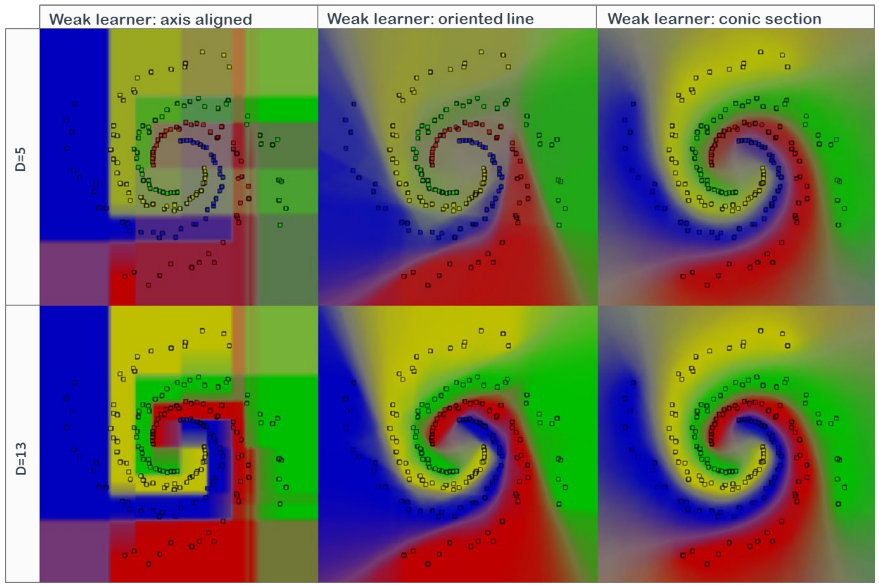
Hyperparameters

Max Features

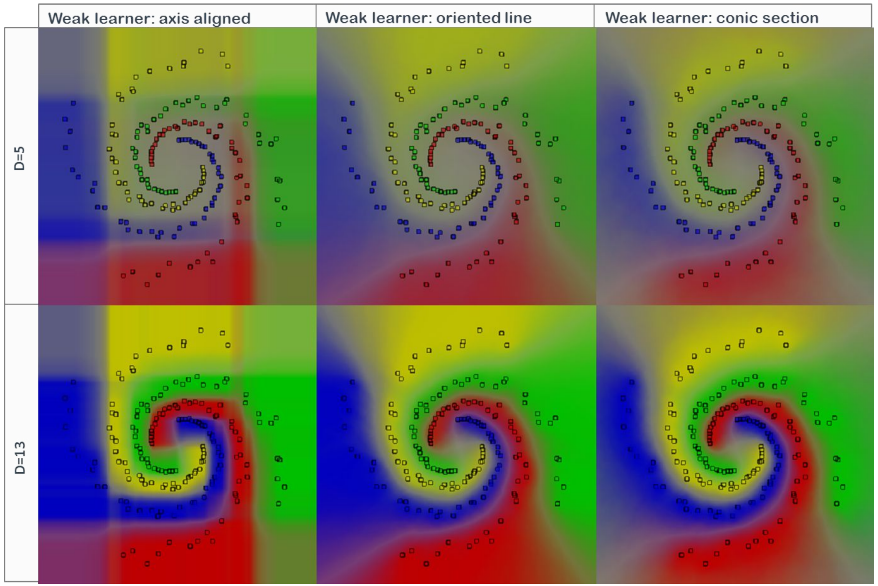
Using less features yields less correlated trees.

Lower confidence in classifications.

500 Max Features



5 Max Features



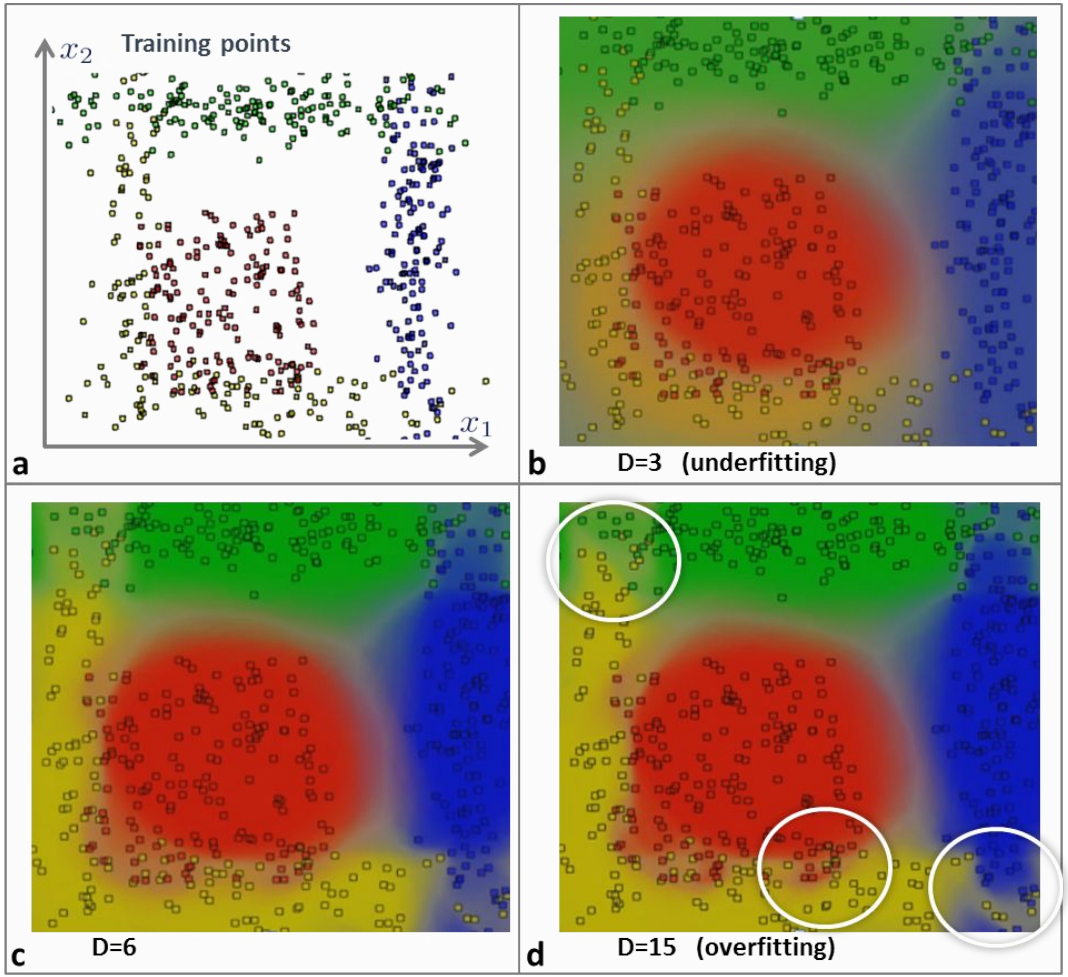
[Criminisi et al 2011]

Hyperparameters

Tree Depth

Higher depth risks overfitting

Lower depth risks underfitting



[Criminisi et al 2011]

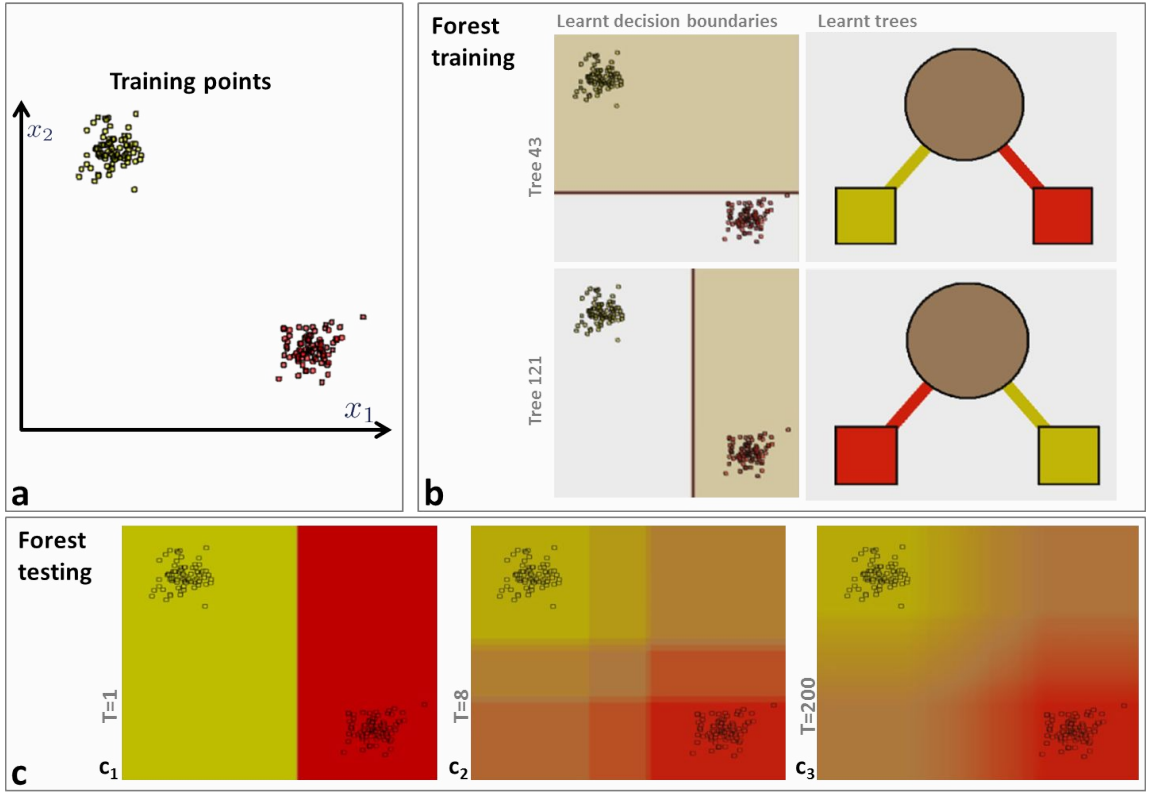
Hyperparameters

Forest Size

Larger forests give higher quality confidence regions.

More trees = More compute

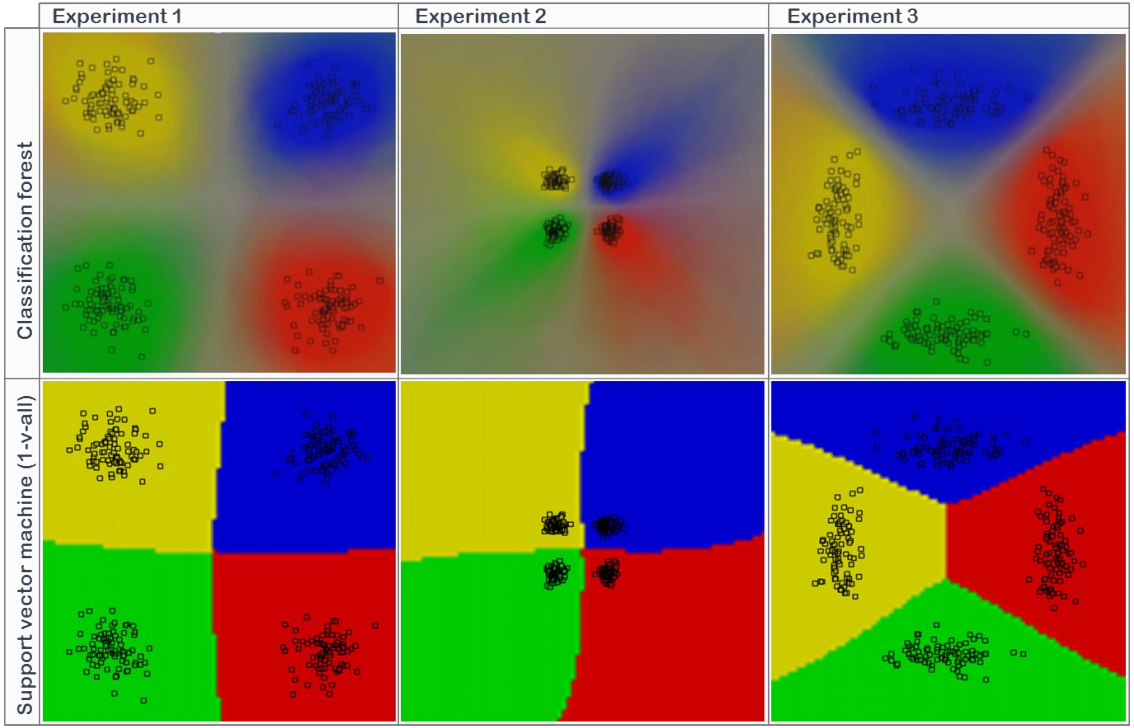
Tend to see diminishing returns after some threshold



[Criminisi et al 2011]

Comparison

SVM vs. RF

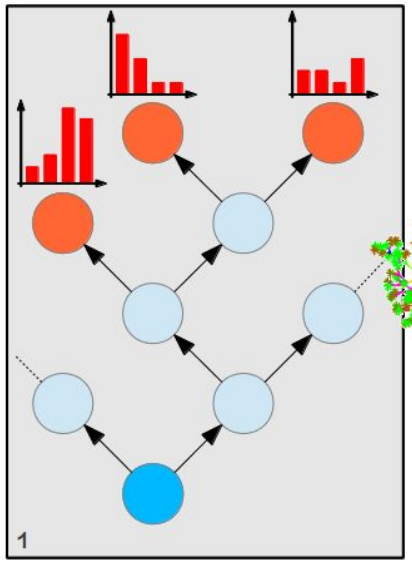


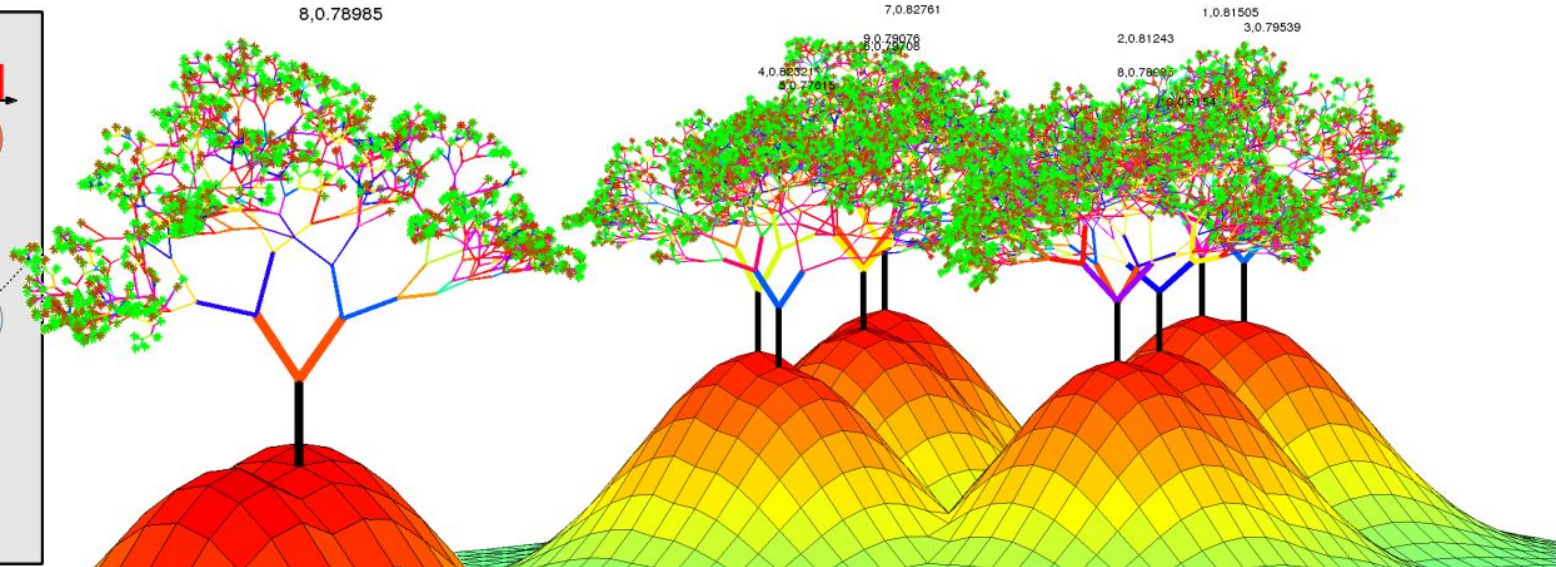
[Criminisi et al 2011]

Continued Learning..

Boosting
Extremely Random Forests
Isolation Forests

Microsoft Research Papers
Scikit-learn documentation





Resources and References

- https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
- http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
- Real-Time Human Pose Recognition in Parts from Single Depth Images, Shotton 2013
- Decision Forests for Classification, Regression, Density Estimation, Manifold Learning, and Semi-Supervised Learning, Criminisi 2011
- Understanding Random Forest, Louppe 2015