Random Forests

CSCI 447/547 MACHINE LEARNING

Outline

- Application Areas
- Benefits
- Construction
- Operation
- Hyperparameters
- Examples

Benefits

- Reduce overfitting
- Training time is less
- Efficient on large databases
- Increases accuracy
- Works well with (lots of) missing data

Construction

- Construct multiple decision trees
  - Choose a random subset of features
  - Choose (or not) a random subset of the training data
- Decision of majority of trees is accepted as the answer

Construction

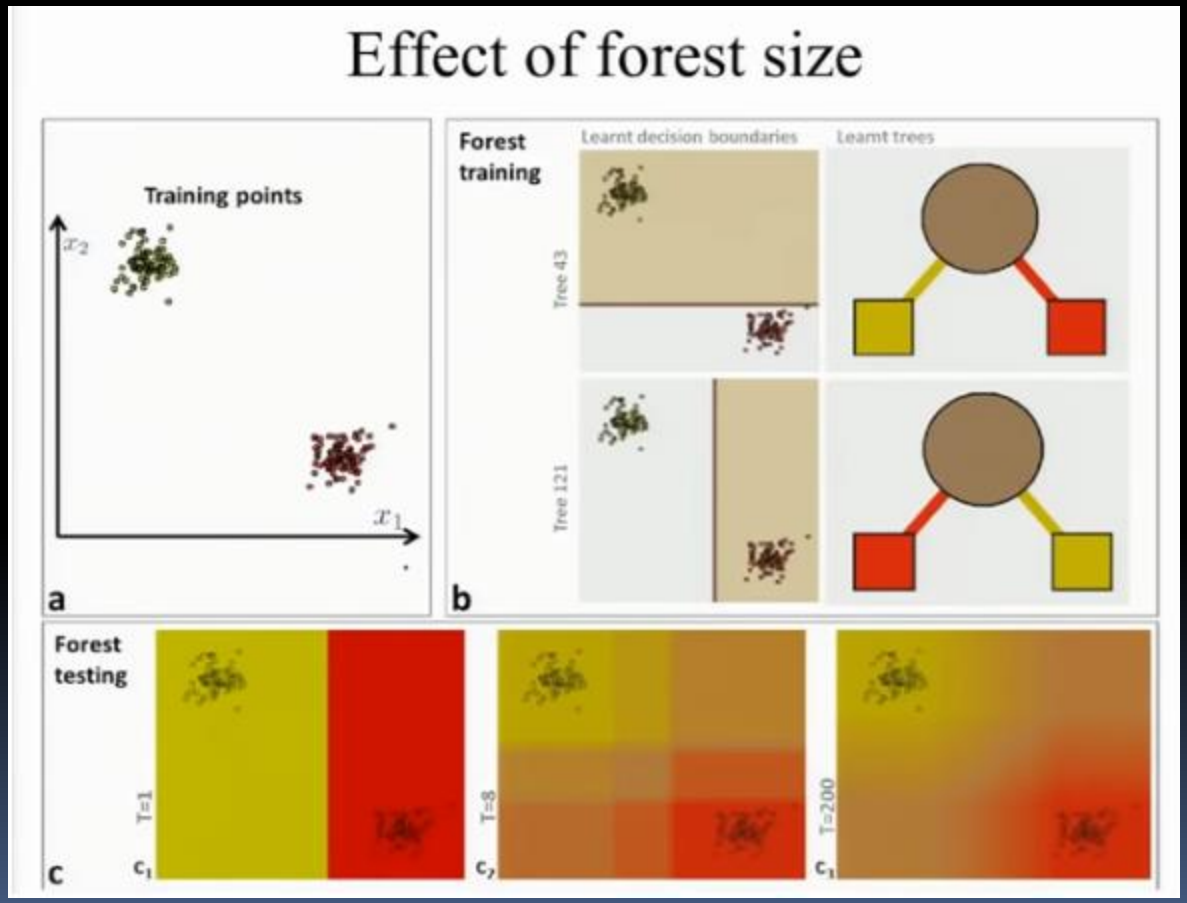
Random Forests algorithm

1. For $b = 1$ to B:
   (a) Draw a bootstrap sample $\mathbf{Z}^*$ of size N from the training data.
   (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
      i. Select m variables at random from the p variables.
      ii. Pick the best variable/split-point among the m.
      iii. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$.

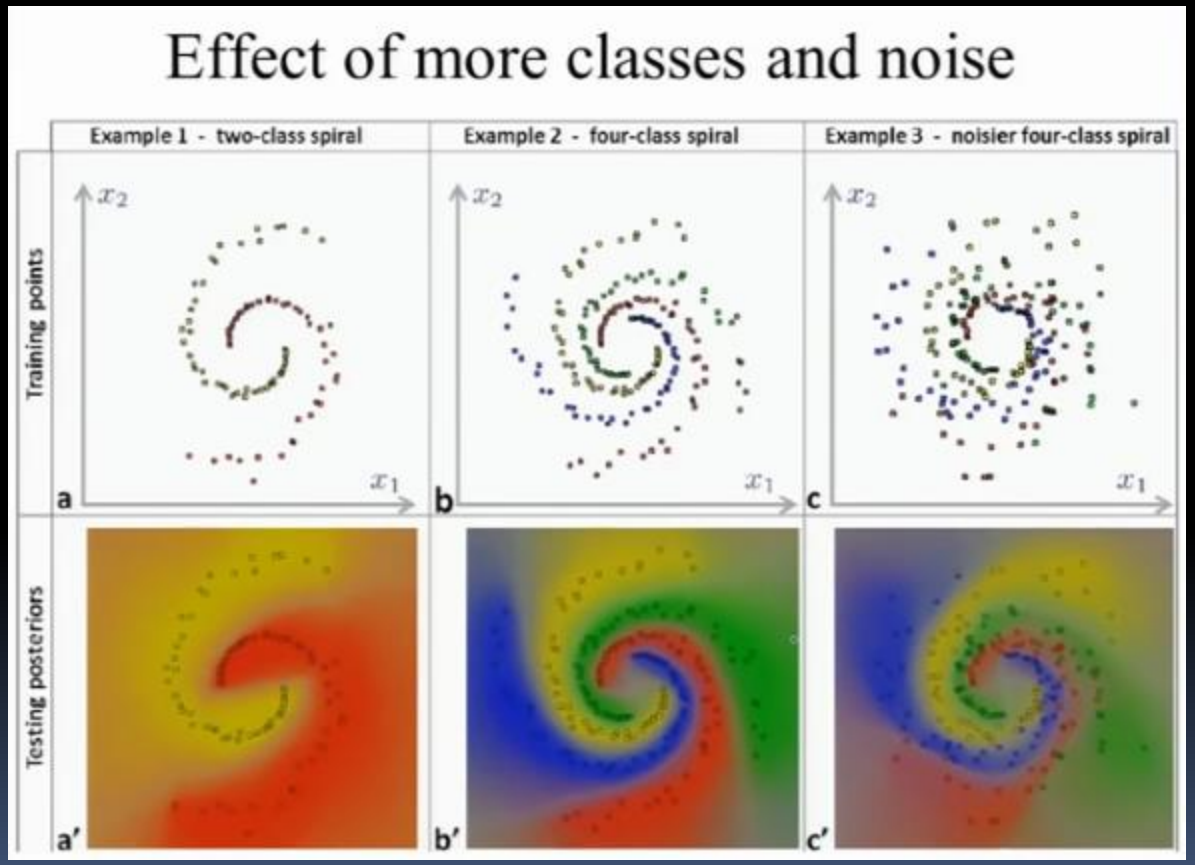
Operation

- End up with some number of trees
- All trees trained independently
  - Possibly in parallel
- During testing, each new example is run through all trees until it reaches a leaf in each
  - Again, possibly in parallel
- Choose maximum vote of trees for final answer, or possibly numeric average

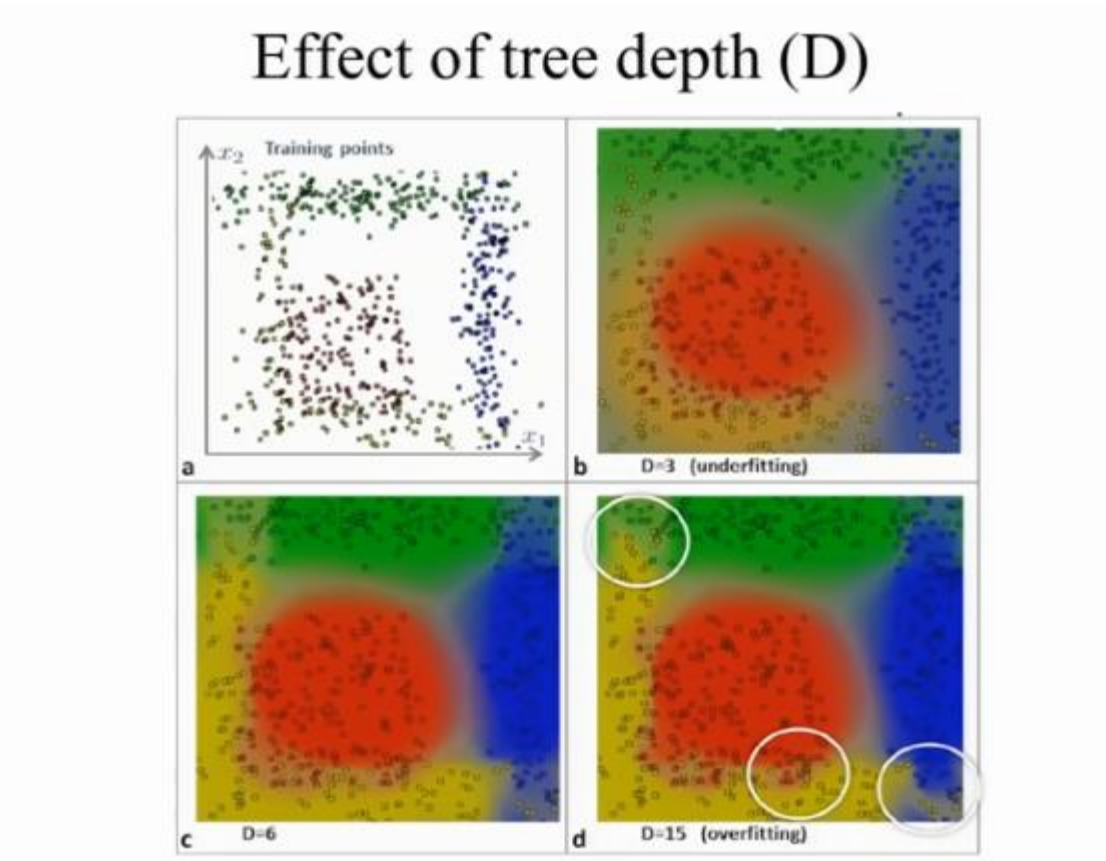
Hyperparameters



Hyperparameters

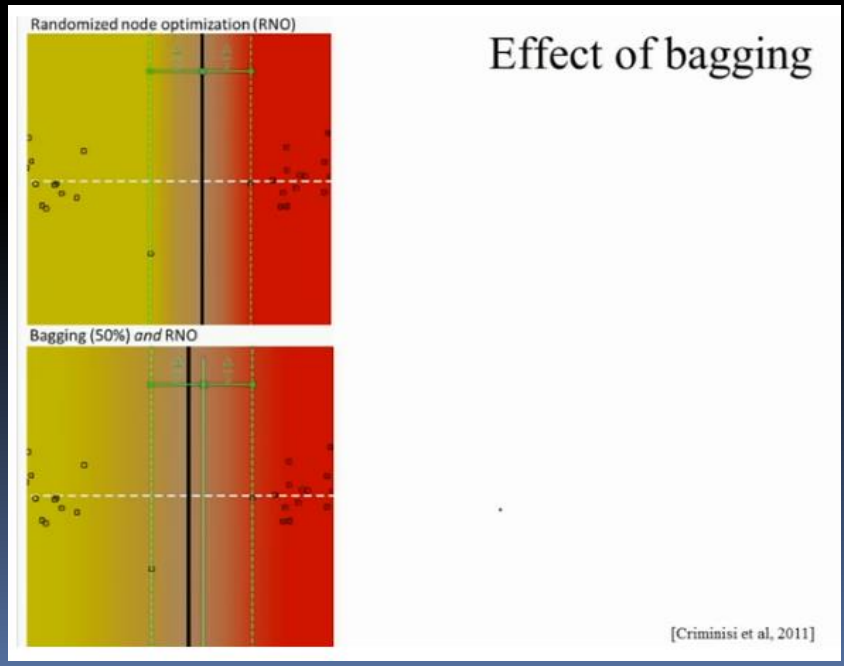


Hyperparameters



Hyperparameters

- Bagging – subsampling the data
- No bagging – max margin (remember SVMs?)
- Bagging – decision boundary may not maximize boundary margin
  - But gives us immunity to outliers



Examples

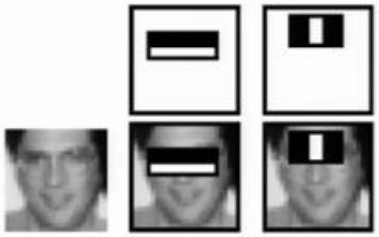
- Training data:
  - 5,000 faces
  - 9,400 non-faces
  - (Viola and Jones, 2001)

Examples

- Use very simple features
  - But when combined into trees, can classify face or not face

Object detection

Idea: Extract simple features from all 24 by 24 pixel patches x_i. E.g., the value of a *two-rectangle feature is the difference between the sum* of the pixels within two rectangular regions. Then compare the level of activation (value of the feature f) with respect to a threshold (theta).

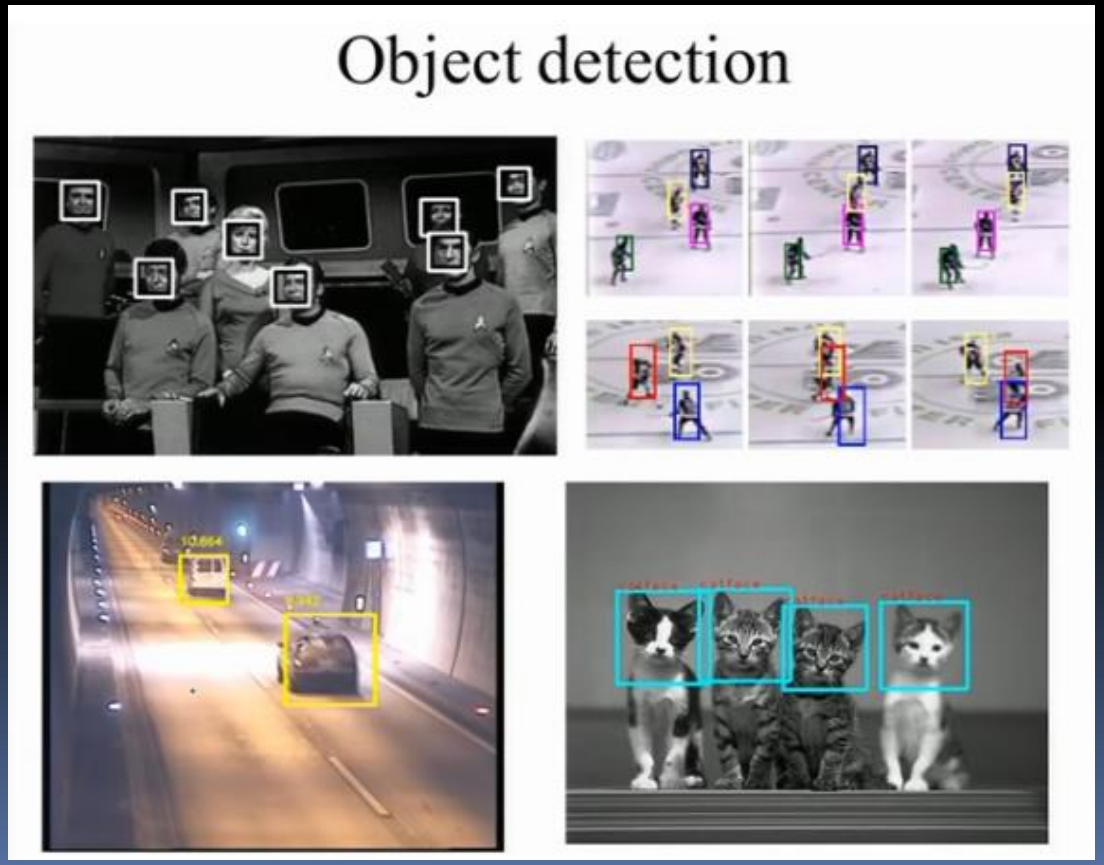


$$h_t(x_i) = \begin{cases} 1 & \text{if } f_t(x_i) > \theta_t \\ 0 & \text{otherwise} \end{cases}$$

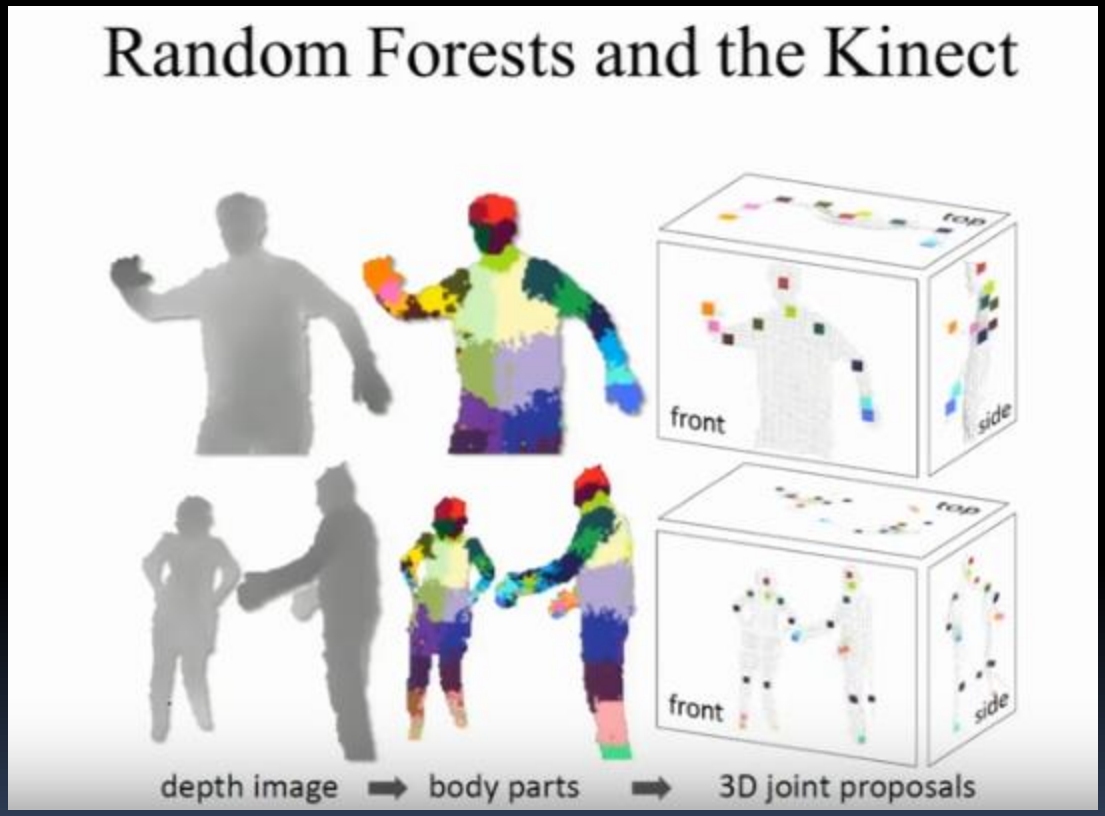
Relevant feature Irrelevant feature

Examples

- Usage beyond just faces:
  - Pedestrians
  - Vehicles
  - And of course, kitties



Examples



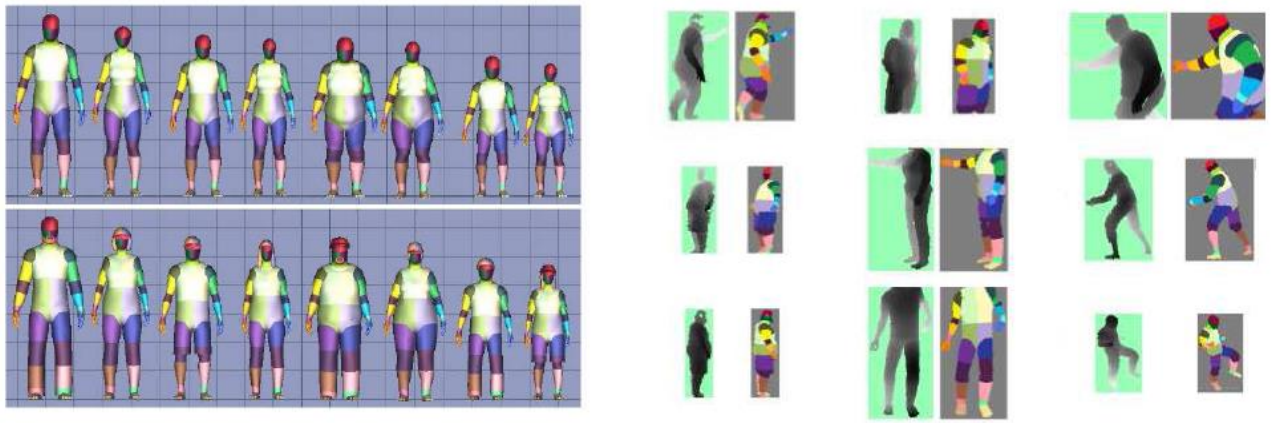
- (Shotton et al, 2011)

Examples

- Remember generating more image data for convolutional networks?

Random Forests and the Kinect

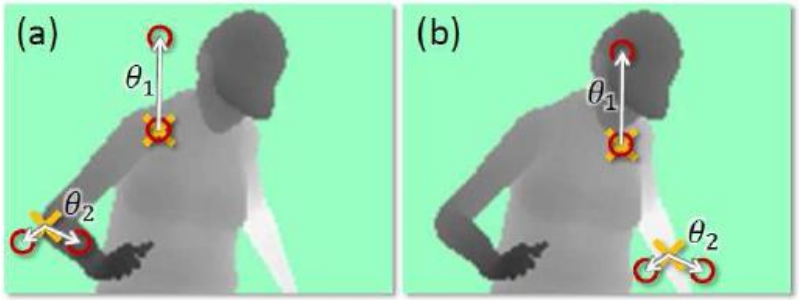
Lesson 1: Use computer graphics to generate plenty of data.



Examples

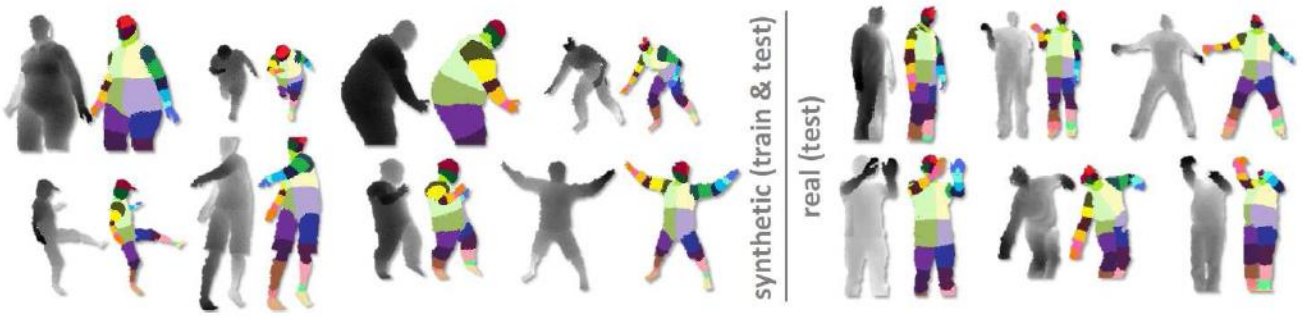
Random Forests and the Kinect

Lesson 2: Use simple depth features within random forests algorithm.



Examples

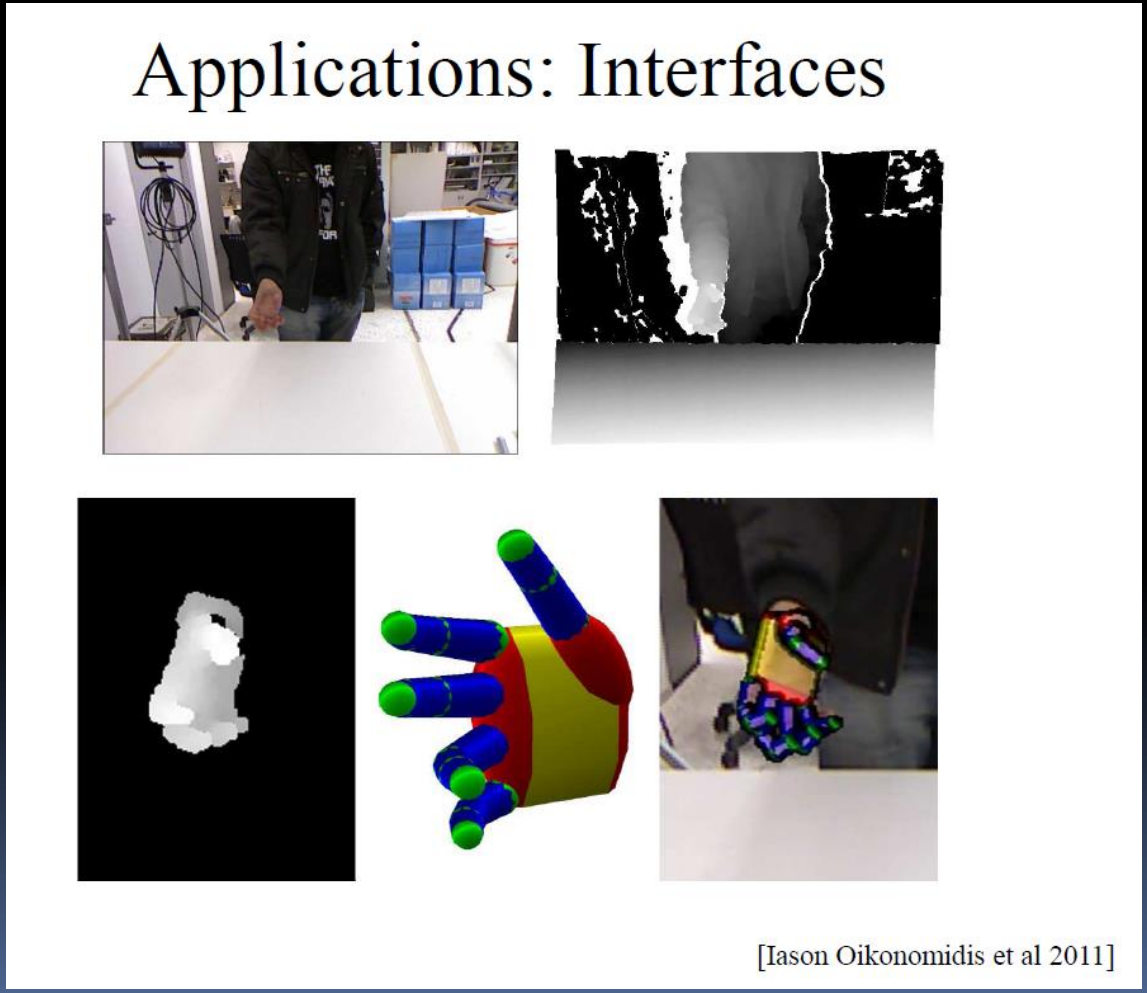
Performance on train and test data



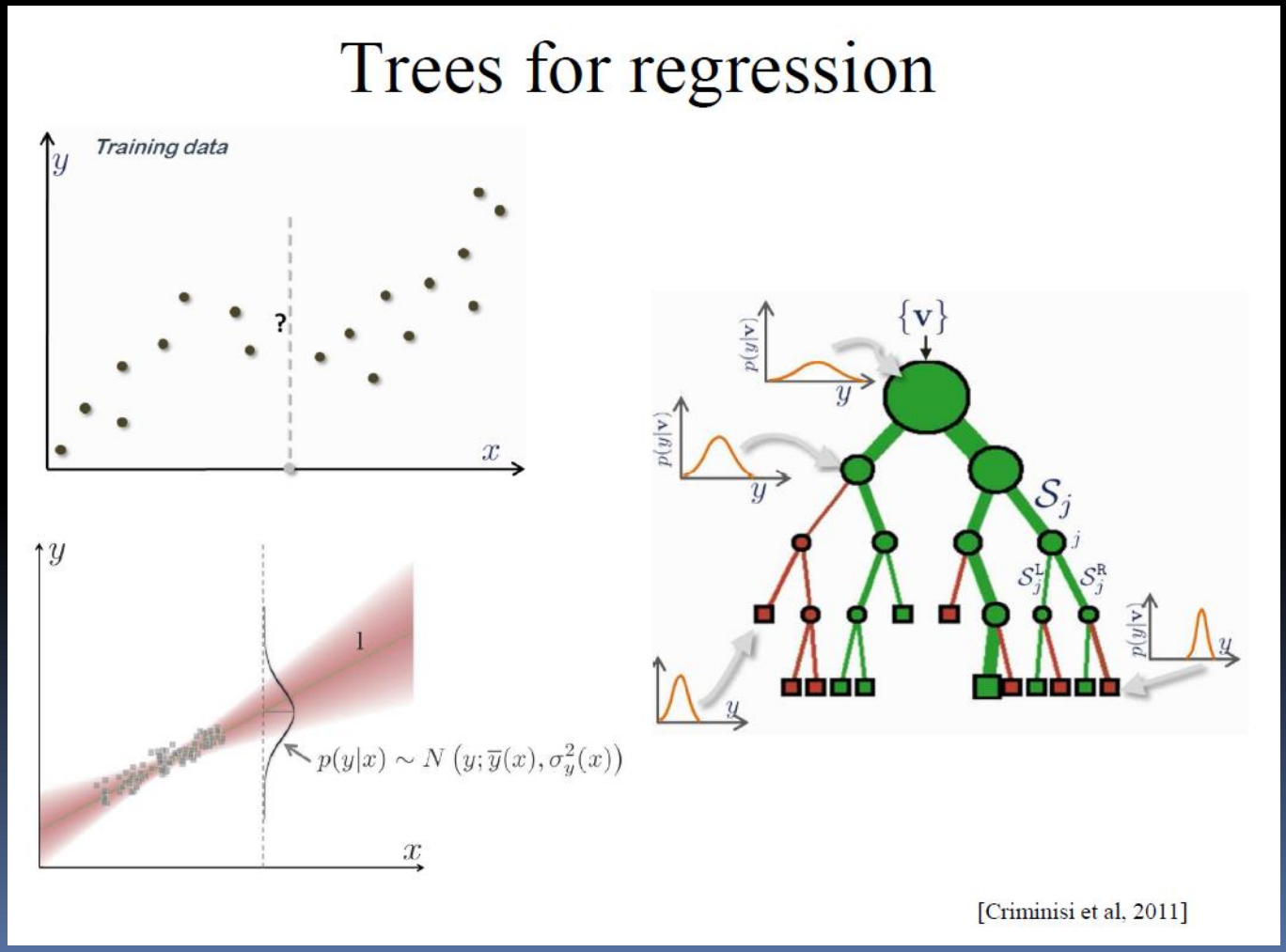
[Jamie Shotton et al 2011]

Examples

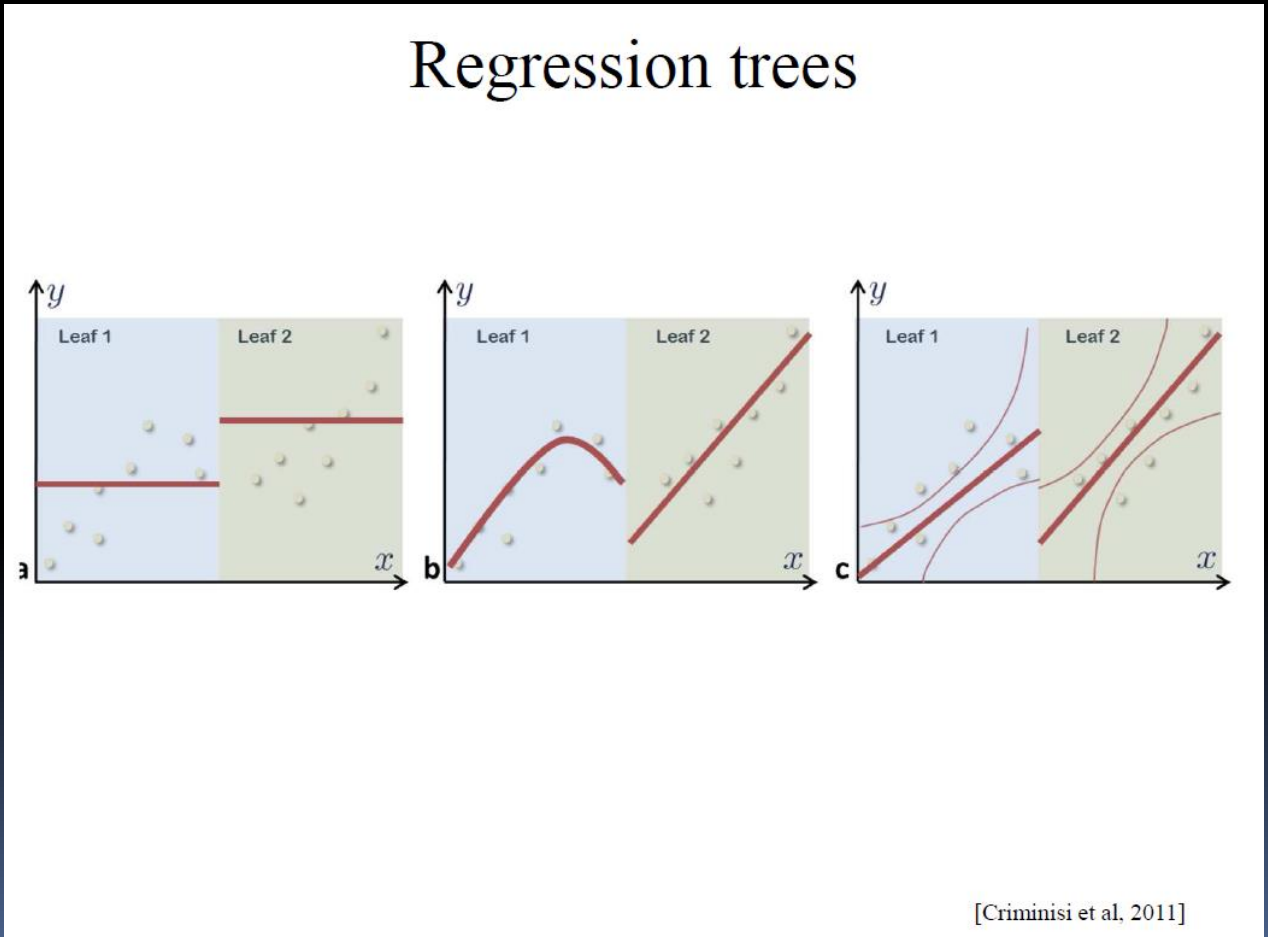
- Applies to areas besides just the Kinect



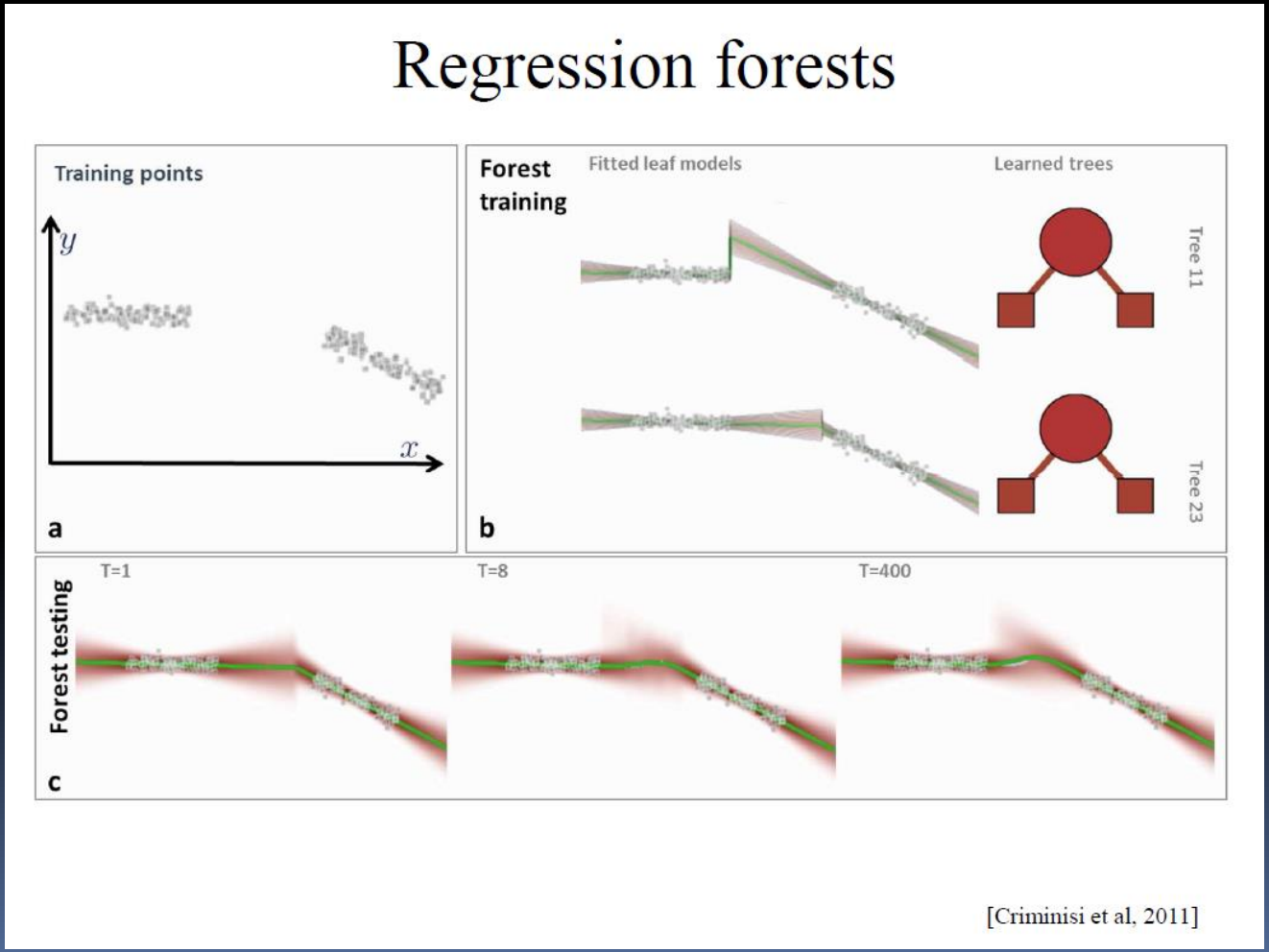
Examples


Trees for regression
Training data
y
x
?
y
x
1
$p(y|x) \sim N\left(y; \overline{y}(x), \sigma_y^2(x)\right)$
{v}
$p(y|\mathbf{v})$
$\mathcal{S}_j$
$\mathcal{S}_j^{\mathrm{L}}$
$\mathcal{S}_j^{\mathrm{R}}$
[Criminisi et al, 2011]

Examples


Regression trees
y
Leaf 1
Leaf 2
x
a
y
Leaf 1
Leaf 2
x
b
y
Leaf 1
Leaf 2
x
c
[Criminisi et al, 2011]

Examples


Regression forests
Training points
y
x
a
Forest training
Fitted leaf models
Learned trees
Tree 11
Tree 23
b
Forest testing
T=1
T=8
T=400
c
[Criminisi et al, 2011]

Summary

- Application Areas
- Benefits
- Construction
- Operation
- Hyperparameters
- Examples