Decisions Trees, Random Forests

Jose Martinez Heras

14/03/2018

Resources

Watch the video of this lecture

https://dlmultimedia.esa.int/download/public/videos/2048/03/005/4803_005_AR_EN.mp4

Watch the practical exercise video

https://dlmultimedia.esa.int/download/public/videos/2048/03/006/4803_006_AR_EN.mp4

Get presentation and additional resources on

https://github.com/jmartinezheras/2018-MachineLearning-Lectures-ESA

Outline for Supervised Learning (2)

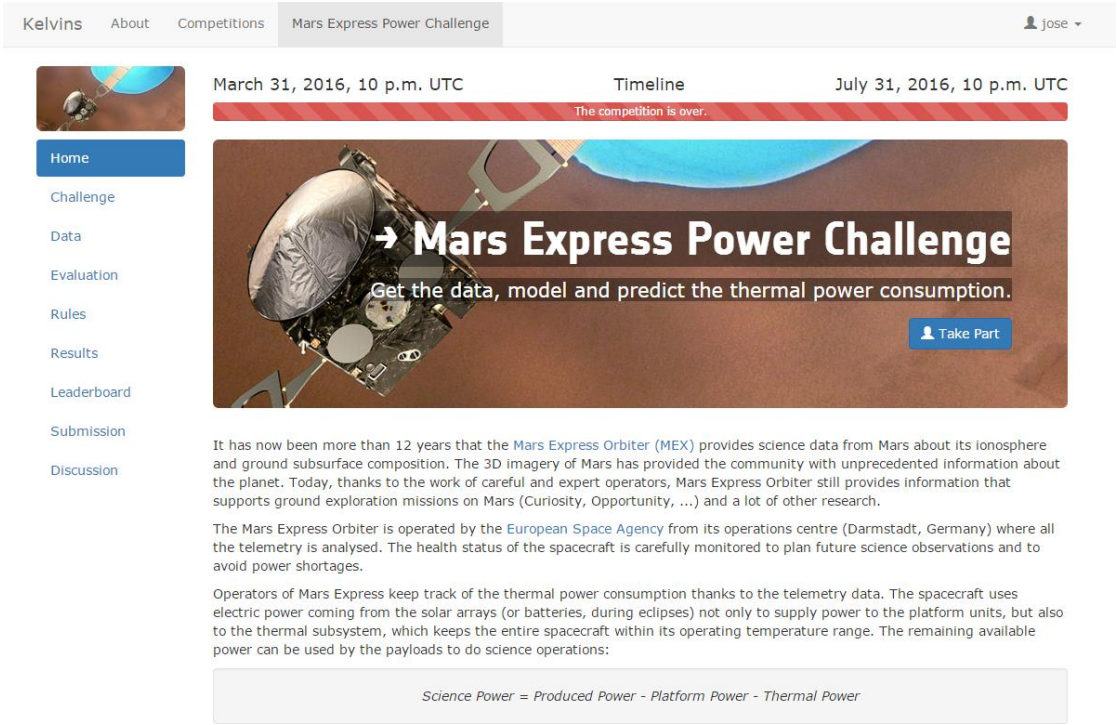
Session 3: Supervised Learning (2)

- Decision Trees
- Ensembles
- Random Forests

- Hands-on

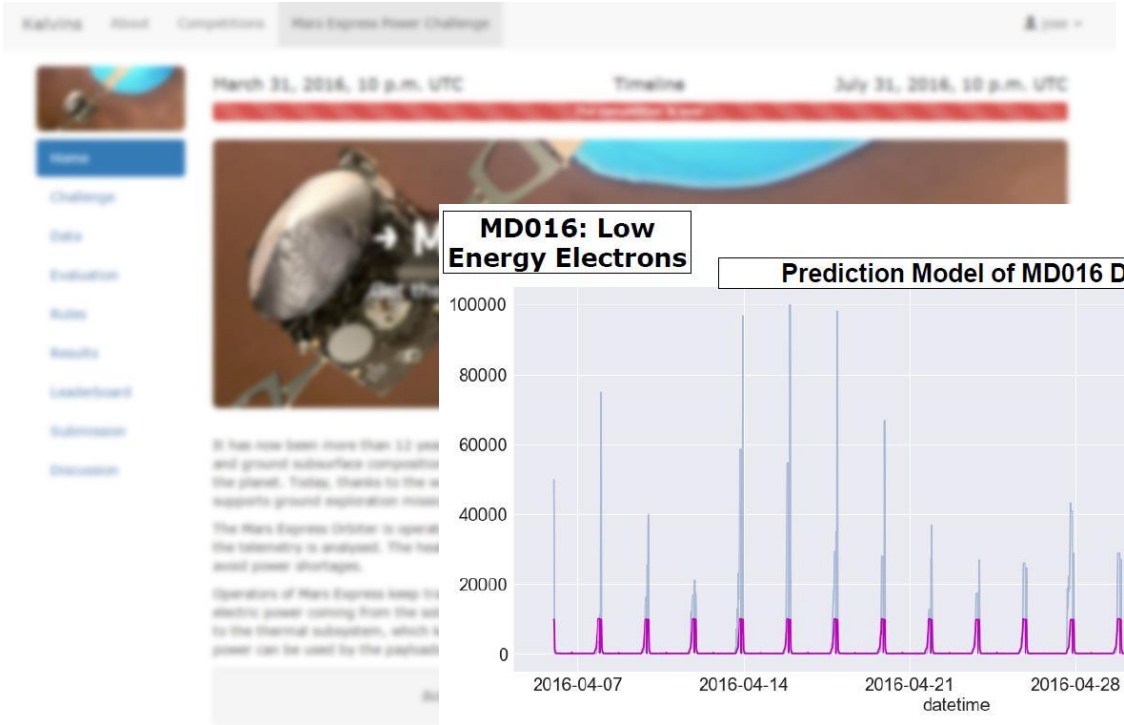
Random Forests

[MEX] Predict Thermal Power Consumption



Random Forests

[MEX] Predict Thermal Power Consumption



MD016: Low Energy Electrons

[Integral] Radiation Belts entry / exit times prediction to increase science return

Importance of features

0.820534935602, altitude
0.0507381593424, rx
0.0476677808188, vx
0.0302844234954, ry
0.0129121849191, vz
0.0106376974037, rz
0.0105089758999, vy
0.00720441182158, PA_rad
0.00315537128504, DEC_rad
0.00272178660137, RA_rad
0.002222400982, ROLL_ANGLE
0.00137769977878, umbra
3.41720505552e-05, penumbra

Random Forests

[MEX] Predict Thermal Power Consumption

[Integral] Radiation Belts entry / exit times prediction to increase science return

Random Forests

Random Forest = Ensemble of Decision Trees

1. Decision Trees
2. Ensembles
3. Random Forests

Decision Trees – remember Iris example?

Iris setosa

Iris versicolor

Iris virginica

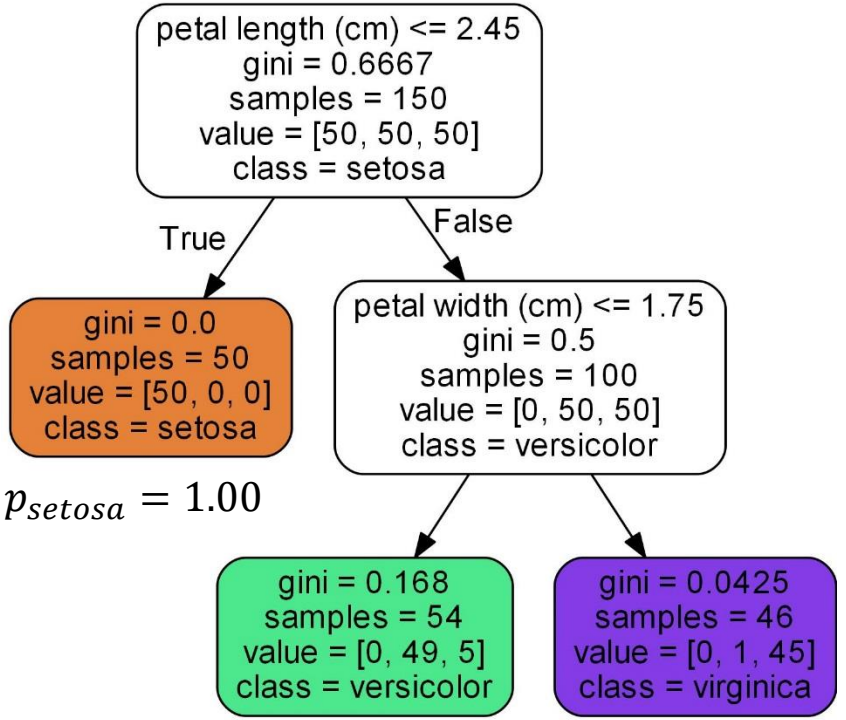
Knowing the sepal and petal length and width, which flower is it?

Pictures from Wikipedia contributors, "Iris flower data set," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Iris_flower_data_set&oldid=824486644 (accessed March 5, 2018).

Decision Trees

Gini is a measurement of purity

It also can provide probabilities



$p_{setosa} = 1.00$

$p_{versicolor} = 0.91$ $p_{virginica} = 0.98$

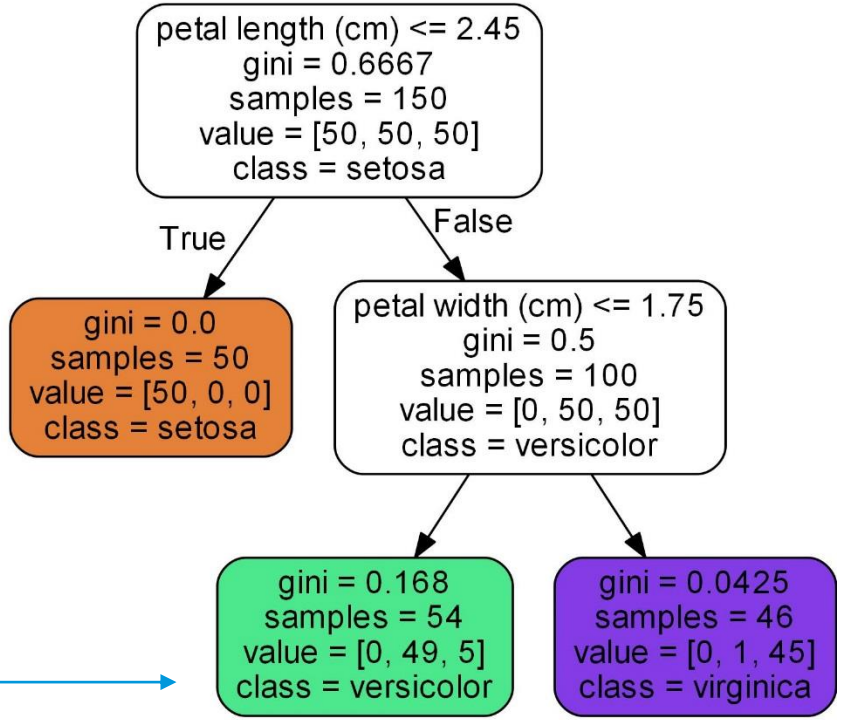
Decision Trees

Gini is a measurement of purity

$$Gini = 1 - \sum_{k=1}^{n} {p_c}^2$$

p_c= probability of each class

$$Gini = 1 - \left(\frac{0}{54}\right)^2 + \left(\frac{49}{54}\right)^2 + \left(\frac{5}{54}\right)^2 = 0.168$$



How do we learn to build Decision Trees?

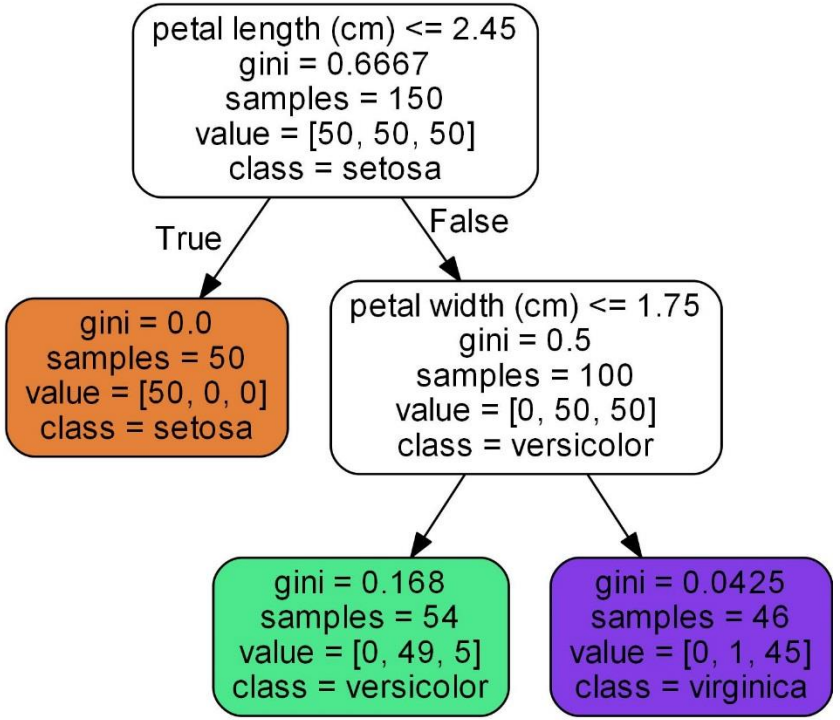
Greedy optimization that minimizes

$$J(f, t_f) = \frac{m_{left}}{m} Gini_{left} + \frac{m_{right}}{m} Gini_{right}$$

f = which feature

t_f = which threshold for feature f

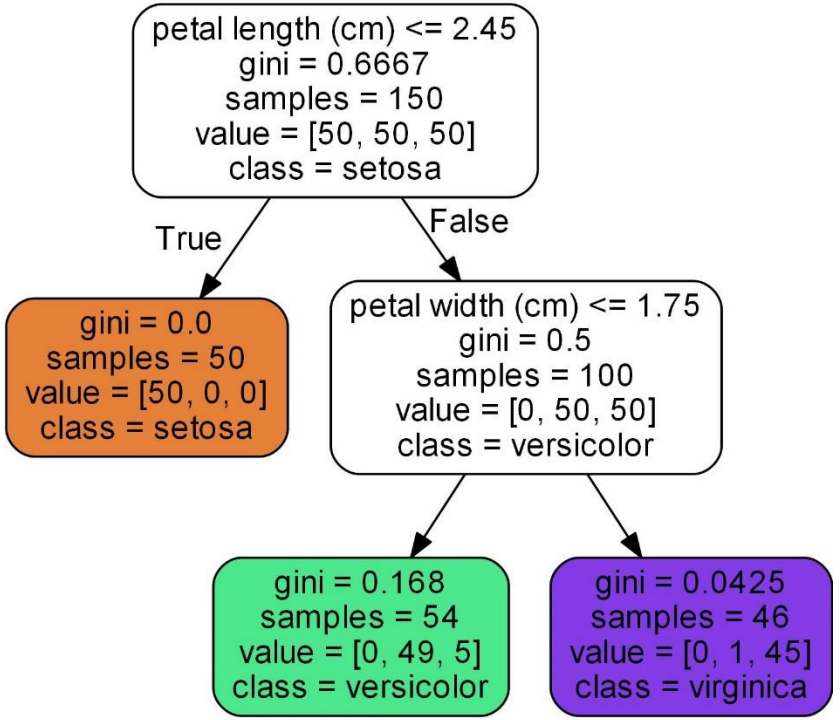
m = number of samples



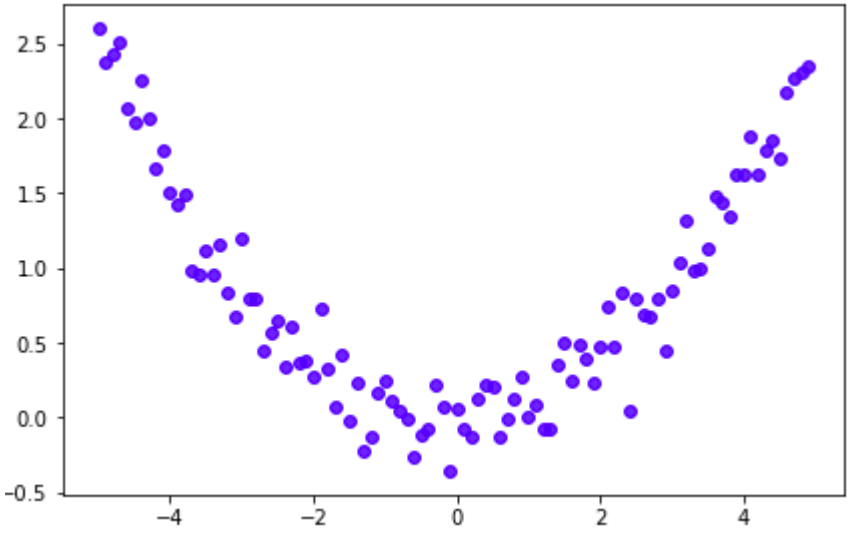
Features Importance

Features Importance

```
sepal length (cm): 0.0
sepal width (cm):  0.0
petal length (cm): 0.562
petal width (cm):  0.438
```

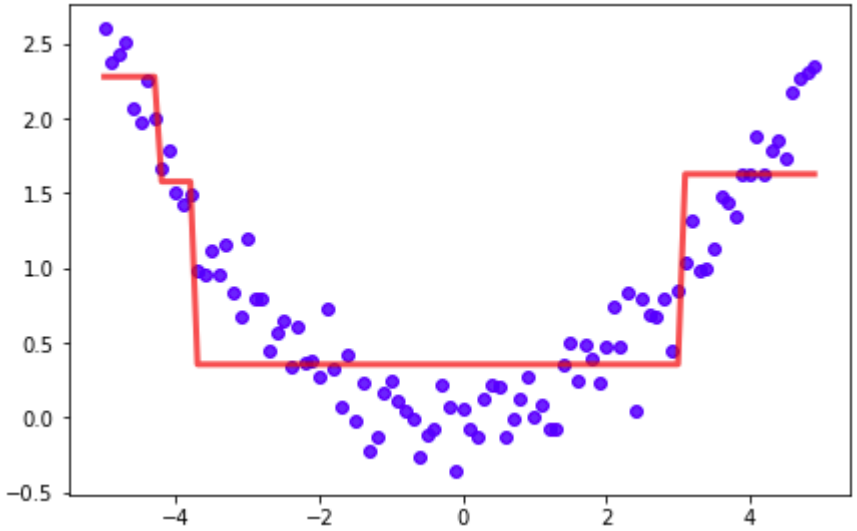


Decision Trees - Regression

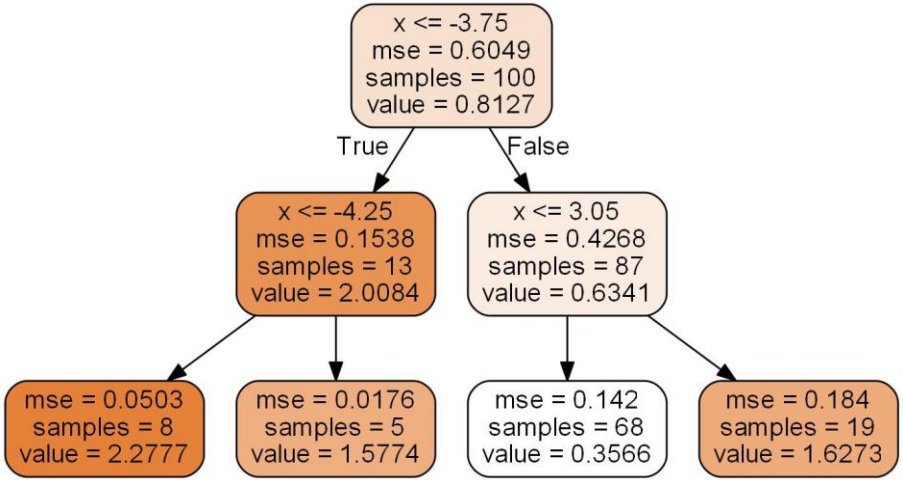


$$y = x^2 + 0.2 \cdot GaussianNoise$$

Decision Trees - Regression



$$y = x^2 + 0.2 \cdot GaussianNoise$$

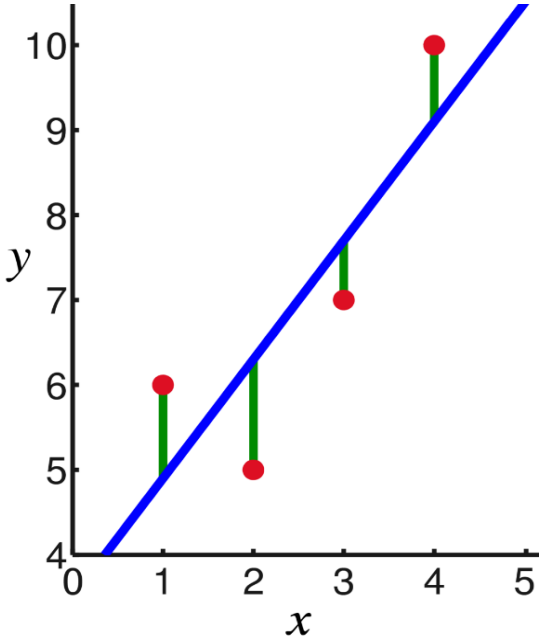


$$J(f, t_f) = \frac{m_{left}}{m} MSE_{left} + \frac{m_{right}}{m} MSE_{right}$$

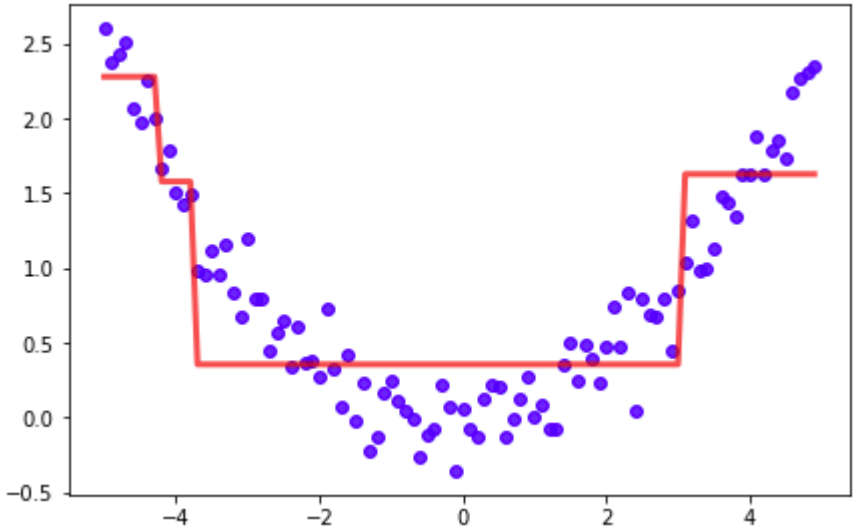
Mean Squared Error

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(Y_i - \hat{Y}_i\right)^2$$

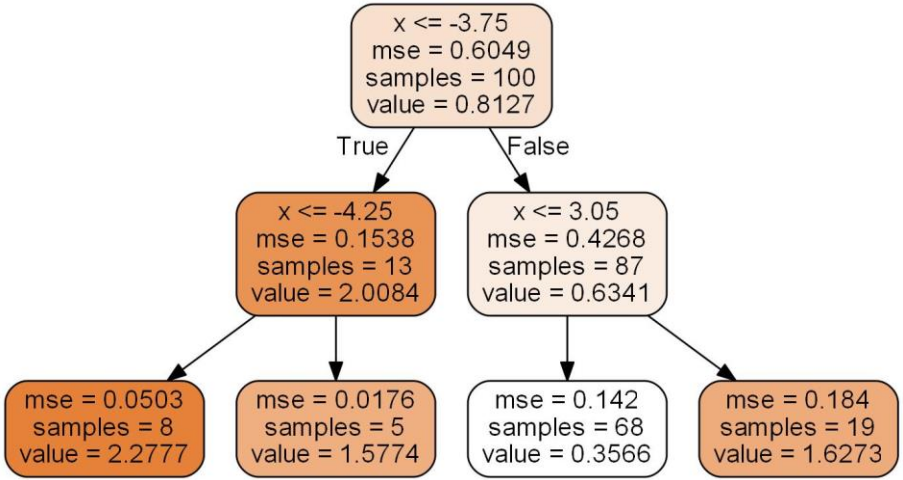
Mean Squared Error



Decision Trees - Regression

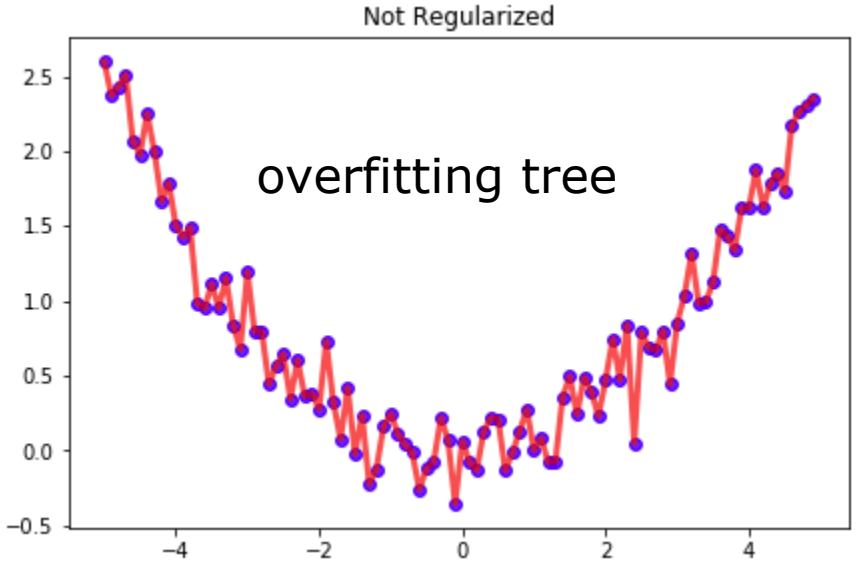


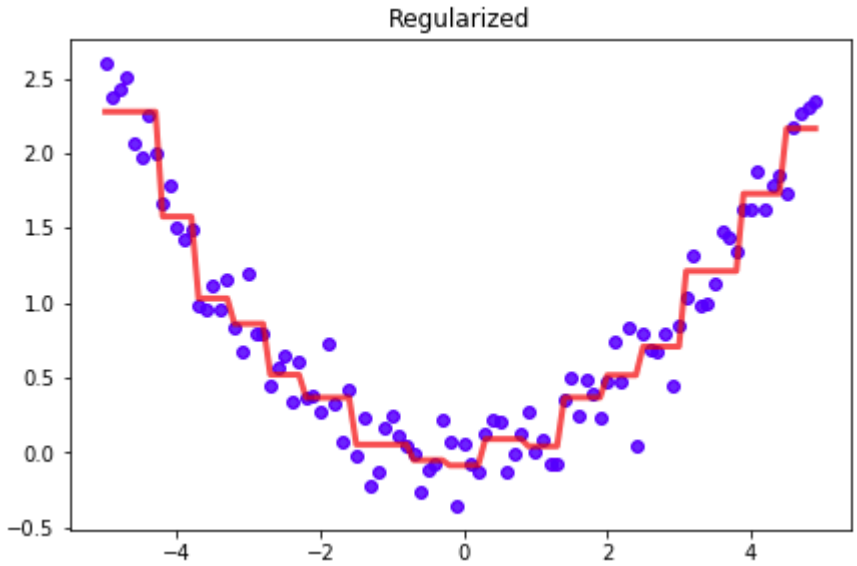
$$y = x^2 + 0.2 \cdot GaussianNoise$$



$$J(f, t_f) = \frac{m_{left}}{m} MSE_{left} + \frac{m_{right}}{m} MSE_{right}$$

Regularization in Decision Trees





For this particular example. Require that each node has ≥ 5% of the samples

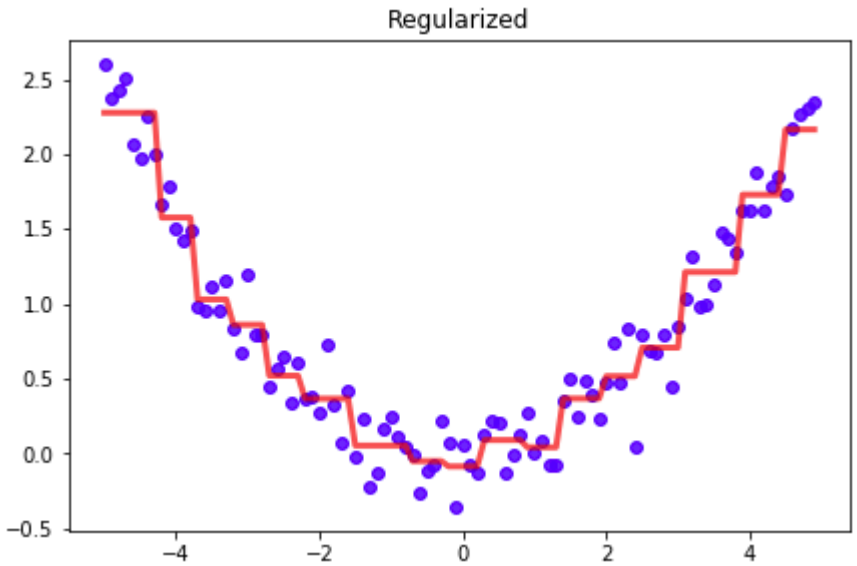
Regularization in Decision Trees

- Maximum depth
- Maximum number of leaf nodes

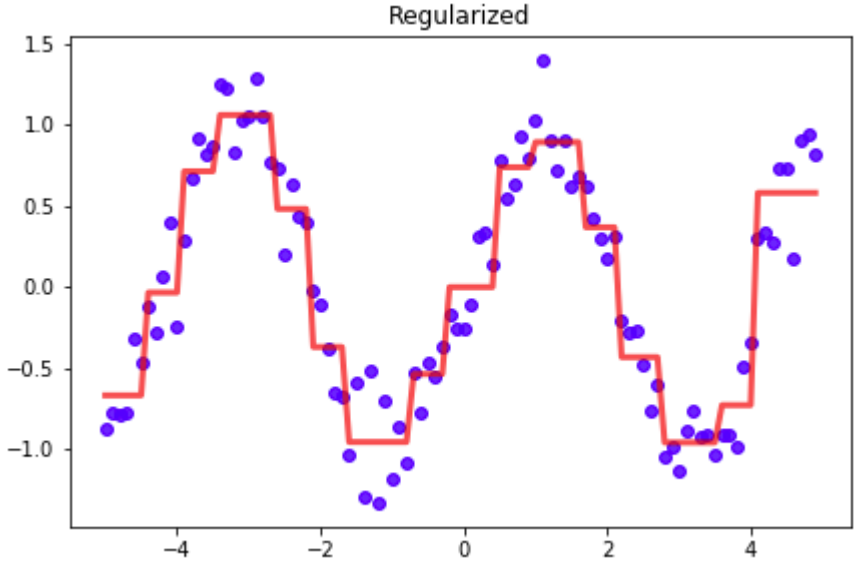
- Minimum number of samples at leaf nodes
- Minimum samples at node must have before it can be split

Decision Trees don’t make assumptions

This is good because it’s not limited by a line, polynomial or any other predefined model



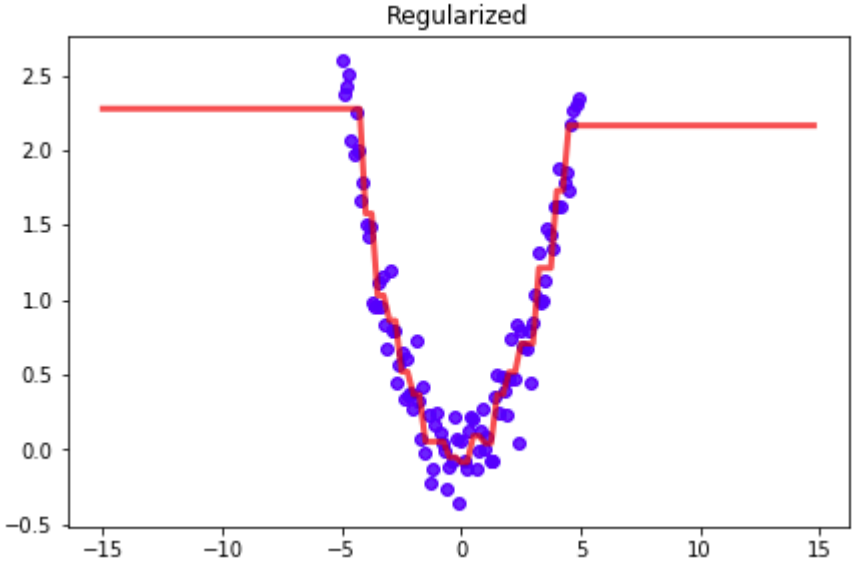
$$y = x^2 + 0.2 \cdot GaussianNoise$$



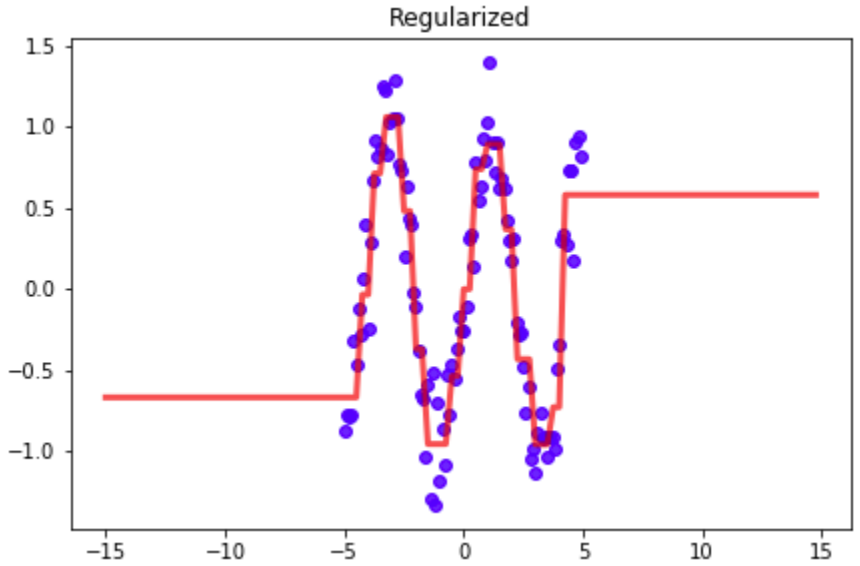
$$y = \sin(1.5 \cdot x) + 0.2 \cdot GaussianNoise$$

Decision Trees don’t make assumptions

But it only works well when trained with representative data



$$y = x^2 + 0.2 \cdot GaussianNoise$$

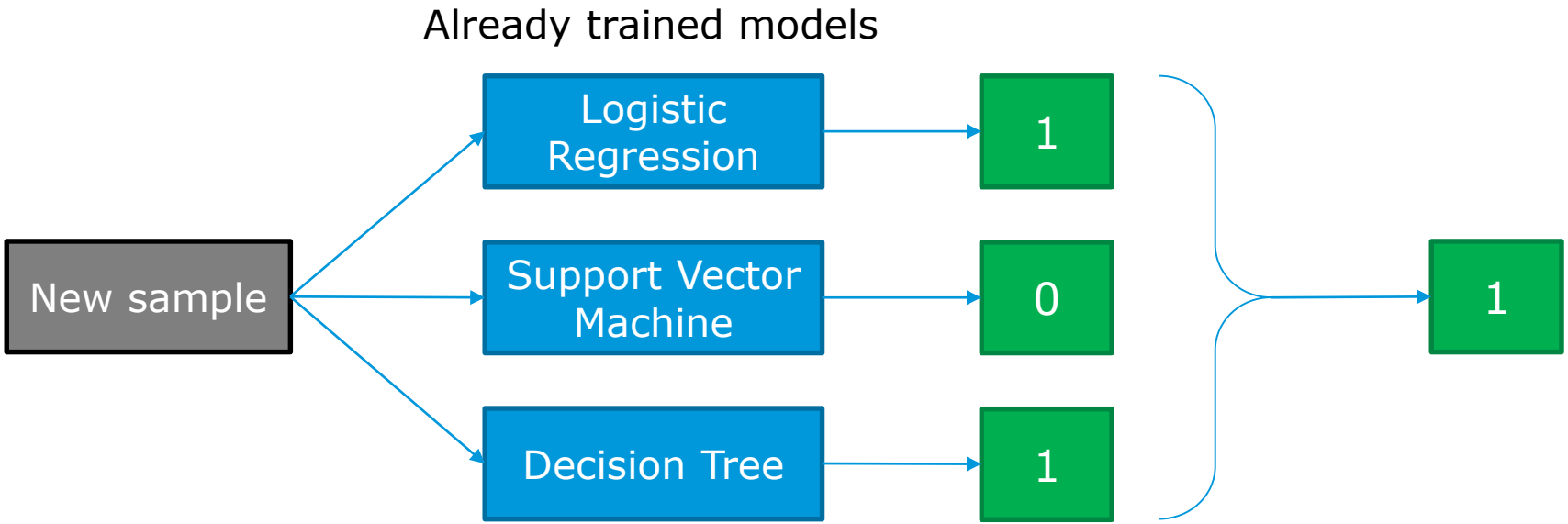


$$y = \sin(1.5 \cdot x) + 0.2 \cdot GaussianNoise$$

Ensembles

- Ensemble = group of Machine Learning models
- Each model in the group produces a different prediction
- Predictions are combined to give a final prediction
- How predictions are **combined**?
  - voting
  - bagging
  - boosting
  - stacking

Ensembles - Voting



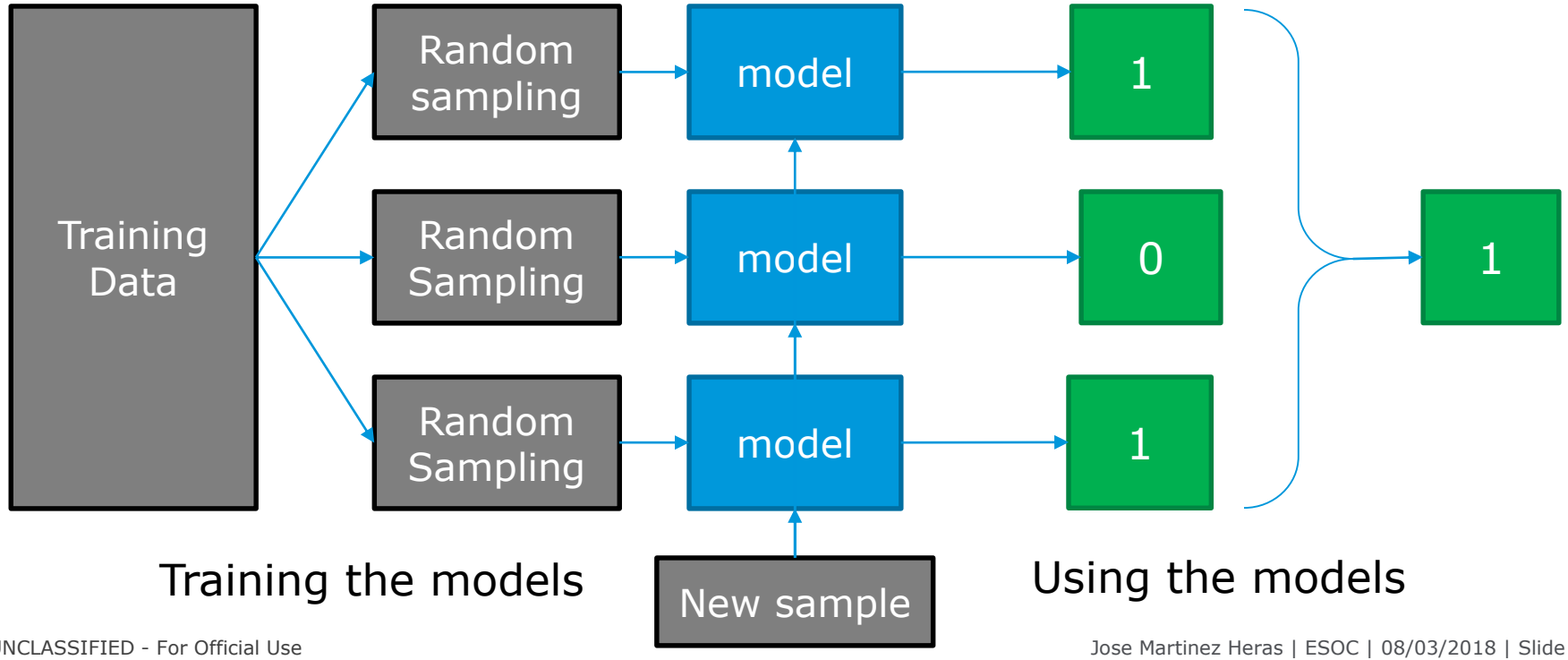
It works better than individual models
As long as models are independent, their errors compensate each other

Ensembles - Bagging

- Voting of several instances of the same model
  - majority voting from several trees
- However
  - all tress would be so similar that their errors will not compensate
- Let’s make Decision Trees different so that we can combine them

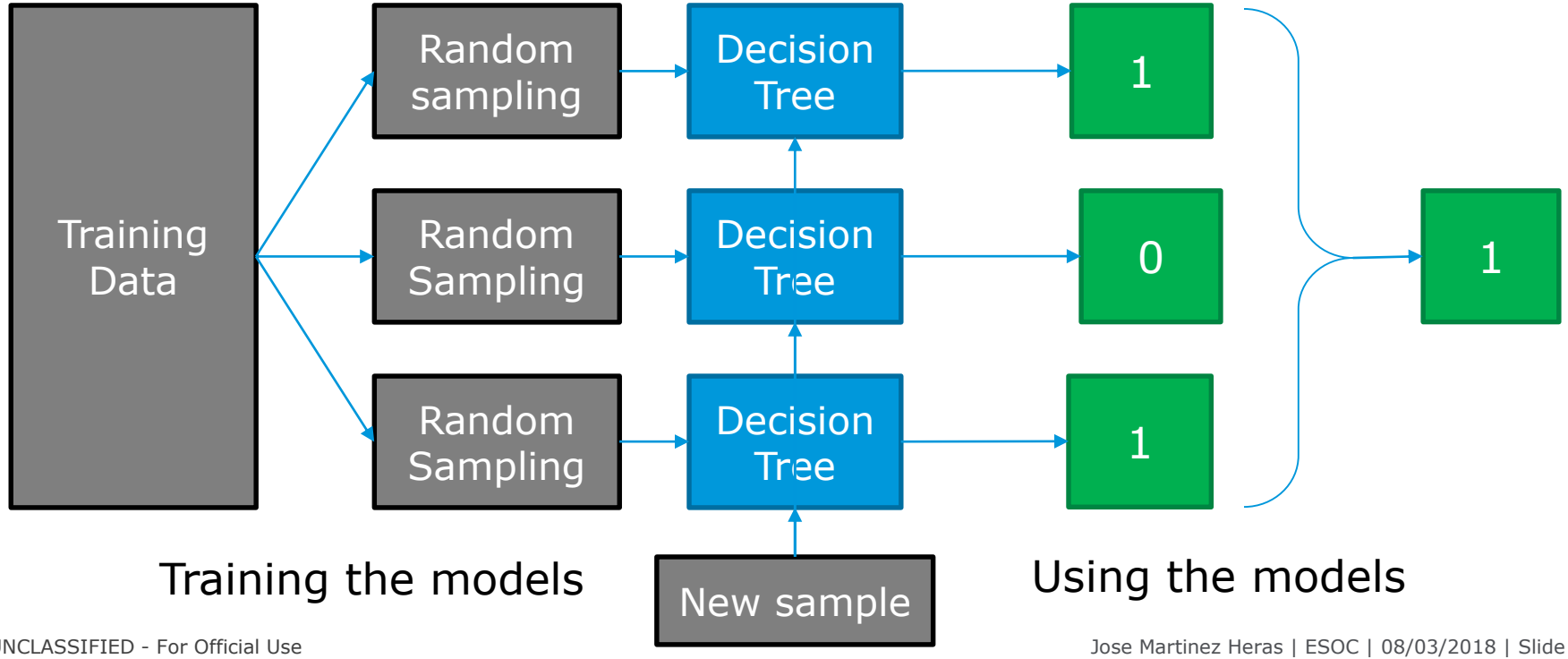
Ensembles - Bagging

Train decision models with **different data / features**



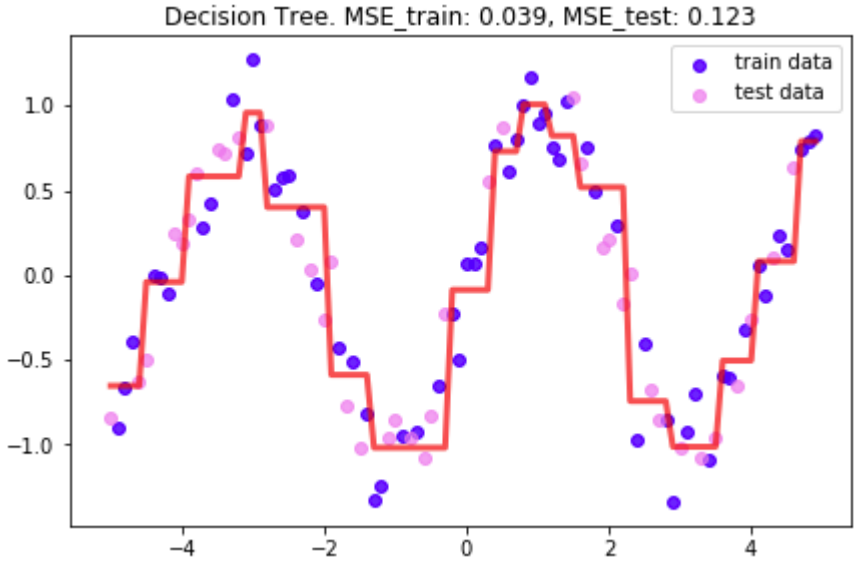
Random Forest: Bagging Ensemble of Trees

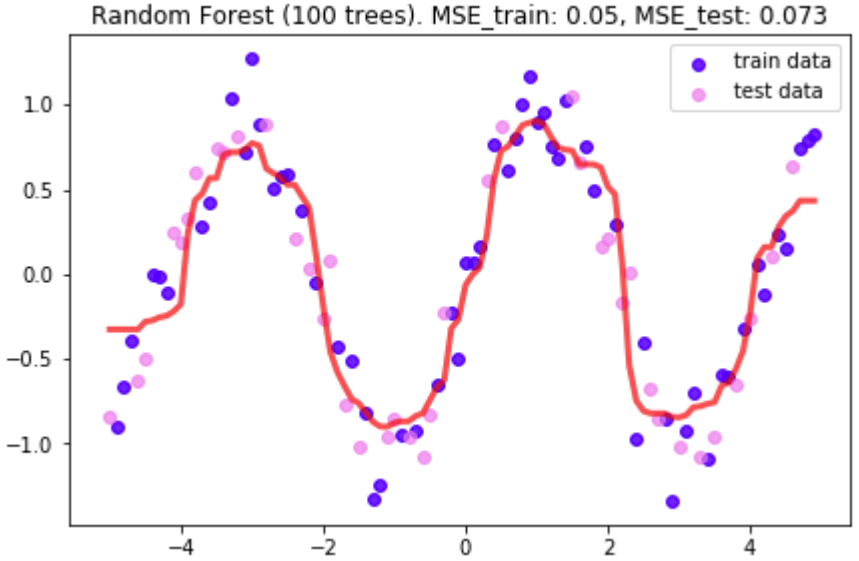
Train decision models with **different data / features**



Random Forests

Random Forest are good for **generalization**





Random Forest provide a **similar training error** (same bias) as an individual tree
Random Forest provide **lower testing error** (lower variance) as an individual tree

How results are combined for Random Forest?

Classification

- Soft-voting
  - Combine the probabilities: highly confident predictions get higher weight

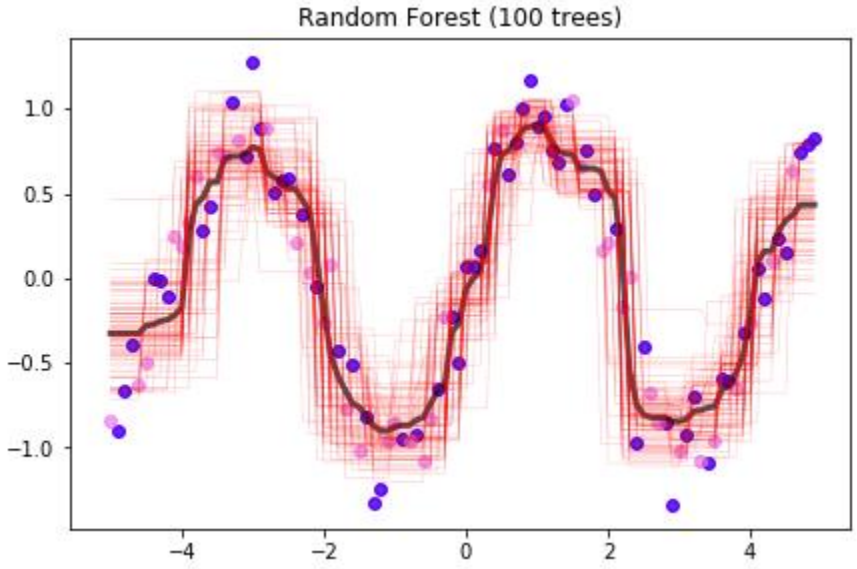
Regression

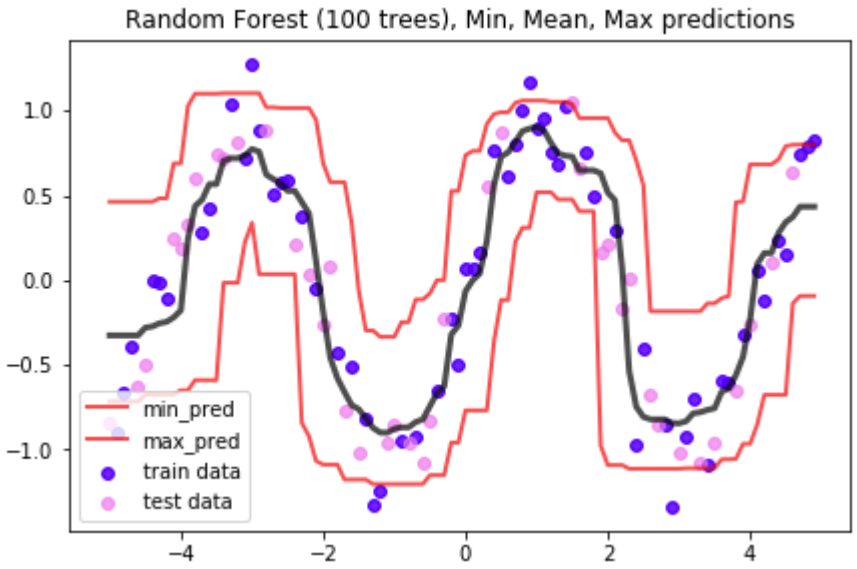
- Mean value

You can also implement your own combination of predictions

How results are combined for Random Forest?

Custom combinations of the Decision Trees Ensembles

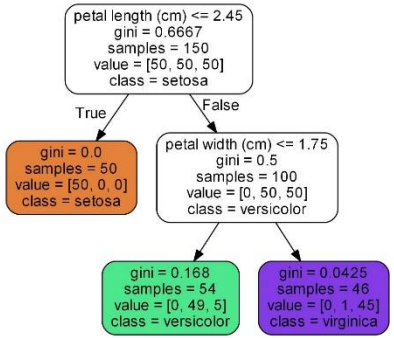




Decision Trees or Random Forest?

Decision Tree

- You are interested in a white-box model
- Understanding is more important than predicting



Random Forest

- You don't mind using a black-box model
- Predicting (generalizing well on new data) is more important than understanding

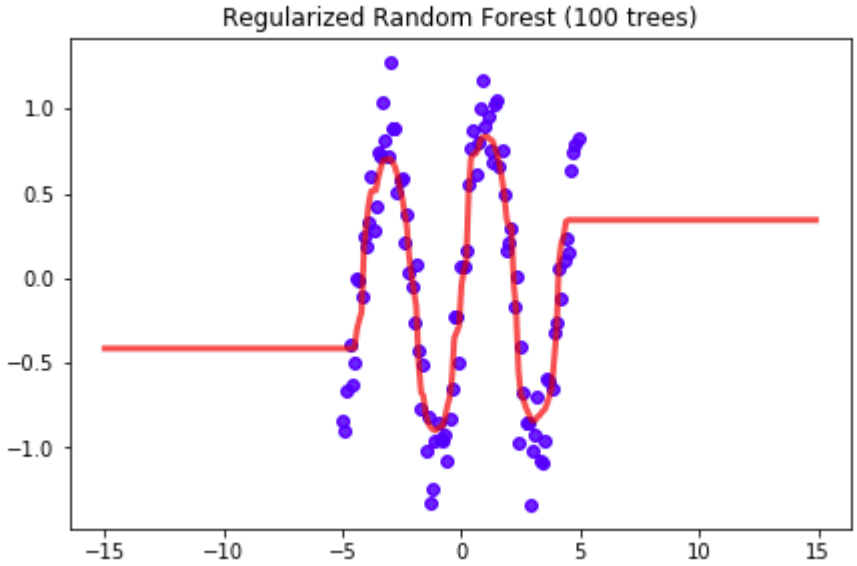
Tree + Random Forest

- Decision Tree for understanding
- Random Forest for predicting

Random Forests

Random Forests gives you **forecasting super-powers**

But **only if** new data is similar to training data



It’s not as bad as I pictured it

Because, in practice, there are many more features

Hands-on

Direct Marketing: predict if a bank term deposit would be (or not) subscribed

Benefits

- Cost reduction
- More deposits
- Less annoyance for clients that are likely not to subscribe the deposit

But before we can work on this Marketing problem,

we need to introduce some concepts

Confusion Matrix / Accuracy / Precision / Recall / F1

How will we evaluate how good our solution is?

		prediction	
		0	**1**
actual	**0**	70	10
	1	15	5

		prediction	
		0	**1**
actual	**0**	TN	FP
	1	FN	TP

Actual Negative (0): 80 cases
Actual Positive (1): 20 cases

TN: True Negative, TP: True Positive
FN: False Negative, FP: False Positive

Confusion Matrix / Accuracy / Precision / Recall / F1

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

		prediction	
		0	**1**
actual	**0**	TN	FP
	1	FN	TP

TN: True Negative, TP: True Positive
FN: False Negative, FP: False Positive

Confusion Matrix / Accuracy / Precision / Recall / F1

misleading

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = 0.75$$

$$precision = \frac{TP}{TP + FP} = 0.33$$ quality

$$recall = \frac{TP}{TP + FN} = 0.25$$ quantity

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} = 0.28$$

quality & quantity

		prediction	
		0	**1**
actual	**0**	70	10
	1	15	5

		prediction	
		0	**1**
actual	**0**	TN	FP
	1	FN	TP

Materials: Slides, Code, Videos

They will be available on the Data Analytics ESA connect community

url: https://connect.esa.int/communities/community/data-analytics

What is next?

March 22th 16:00 – Press Room

Session 4: Supervised Learning (3)

- Neural Networks
- Deep Learning

- Hands on

Resources

Watch the video of this lecture

https://dlmultimedia.esa.int/download/public/videos/2048/03/005/4803_005_AR_EN.mp4

Watch the practical exercise video

https://dlmultimedia.esa.int/download/public/videos/2048/03/006/4803_006_AR_EN.mp4

Get presentation and additional resources on

https://github.com/jmartinezheras/2018-MachineLearning-Lectures-ESA

Thank you

Data Analytics Team for Operations (DATO)

Jose Martinez Heras

LinkedIn: https://www.linkedin.com/in/josemartinezheras/