UBS

Global Research

January 2016

# Machine learning

## *Trees and more trees*

**Claire Jones**
Quantitative Analyst
Tel: +44 20 7568 1873
Claire-c.Jonesl@ubs.com

**Nick Baltas**
Quantitative Analyst
Tel: +44 20 7568 3072
Nick.Baltas@ubs.com

# Table of Contents

# Introduction

- This presentation focuses on classification problems.
- We will consider decision trees and two machine learning algorithms based on them: random forests and adaBoost.
- The 2014 paper, "Do we need hundreds of classifiers to solve real world classification problems?" (2014), Fernandez-Delgado et al, Journal of Machine Learning, compares the effectiveness of 179 different machine learning approaches on 121 different data sets and rank the classifiers.
- Random forests and adaBoost perform extremely well:

| Classifier | Average Accuracy | Rank (out of 179) |
|---|---|---|
| Random Forest | 81.9% | 5 / 179 |
| AdaBoost | 79.1% | 20 / 179 |
| Decision Tree | 73.5% | 108 / 179 |

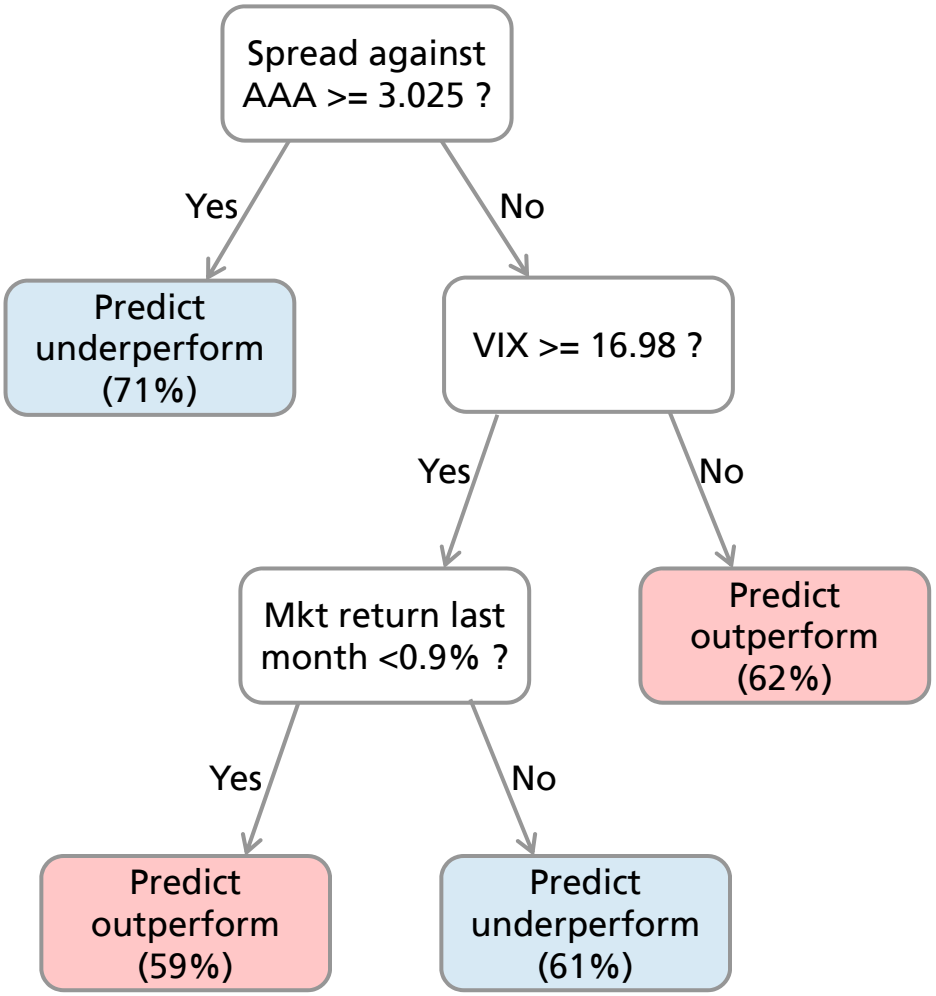
Source: "Do we need hundreds of classifiers to solve real world classification problems?" (2014), Fernandez-Delgado et al, Journal of Machine Learning, Extract from Tables 2 & 3 ,

Section 1

# Decision Trees

# How to use a tree

- The output looks like a flow chart.
- To make a prediction for today, you work your way down the tree until you reach a "leaf".
- The leaf gives you a forecast class and an attached probability e.g. underperform with a probability of 71%.




Example of a classification tree

Source: UBS Quants

# Pros and cons of decision trees

**Pros**

- Completely transparent
- Easy to interpret and communicate
- Robust to outliers / non-normal inputs and works with categorical and continuous data inputs
- Does not rely on linear relationships

**Cons**

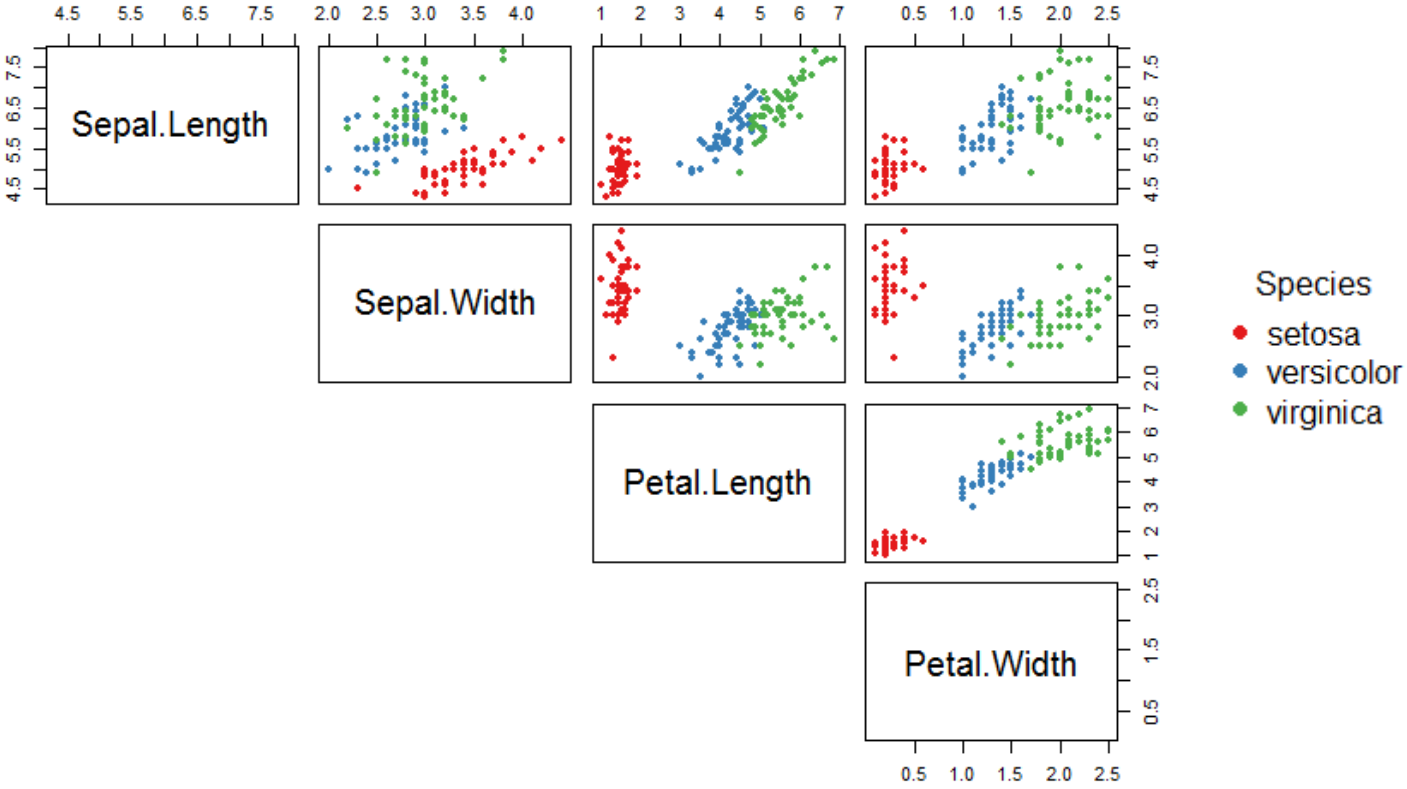
- Decision trees are often unstable i.e. the tree can change completely with the addition of an extra data-point
- Doesn't cope well with severe class imbalances
- Doesn't forecast well for inputs outside the range of its training set
- Sometimes over-fit

# Our example data

- Iris is a data-set about irises. We have data about sepal and petal lengths and widths, and we want to create a model to classify the species of iris.

| Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|---|---|---|---|---|
| 5.9 | 3.0 | 5.1 | 1.8 | viriginica |
| 5.6 | 2.9 | 3.6 | 1.3 | versicolor |
| 5.0 | 3.6 | 1.4 | 0.2 | setosa |
| 5.1 | 3.7 | 1.5 | 0.4 | setosa |
| etc | | | | |

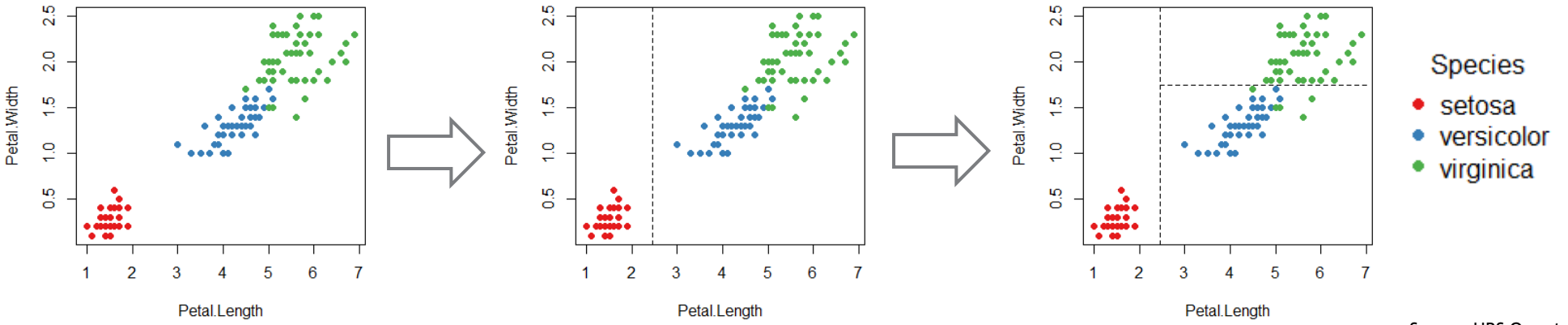
Source: UBS Quants




Source: UBS Quants

# How do we grow a tree?

- We start with all of our data.
- If your data-set is finite then for each feature there are a finite number of possible thresholds, so we can test every possible combination of feature and threshold and choose the one which divides the data "most cleanly" into two homogenous sub-groups.
- Then we repeat the process with each of those two sub-groups separately and so on recursively.




Source: UBS Quants

# How do we choose the feature and threshold at each node?(1)

- Start by defining the impurity of the data with the Gini coefficient:

$$Impurity = 1 - \sum_{i=1}^{\#\ of\ classes} {p_i}^2$$

where pi is the proportion of the data from class i. In this example, we have three classes for each of the three species of iris.

  - This will be zero if all the group is from a single class and bigger if the data is more mixed.
  - The worst case scenario is if the group has equal numbers of points from each class.

- If we cut the data into two sub-groups then the overall impurity is a weighted sum of the impurity in each sub-group:

$Overall\ impurity =$ % of data in 1st subgroup x impurity in 1st subgroup +
% of data in 2nd subgroup x impurity in 2nd subgroup

- We want to choose the cut which minimises the overall impurity i.e. creates two sub-groups which are almost exclusively a single class.

# How do we choose the feature and threshold at each node?(2)

- As an example, here is a table showing the breakdown of iris species where the sepal length is longer or shorter than 5 cm:

| | Species of iris | | | |
|---|---|---|---|---|
| | Setosa | Versicolor | Virginica | Total |
| Overall | 50 | 50 | 50 | 150 |
| Sepal.Length ≤ 5 | 28 | 3 | 1 | 32 |
| Sepal.Length > 5 | 22 | 47 | 49 | 118 |

Source: UBS Quants

$$Overall\ impurity\ before\ cut = 1 - (\frac{50}{150})^2 - (\frac{50}{150})^2 - (\frac{50}{150})^2 = \mathbf{0.66}$$

$$Impurity\ in\ short\ sepal\ subgroup = 1 - (\frac{28}{32})^2 - (\frac{3}{32})^2 - (\frac{1}{32})^2 = \mathbf{0.22}$$

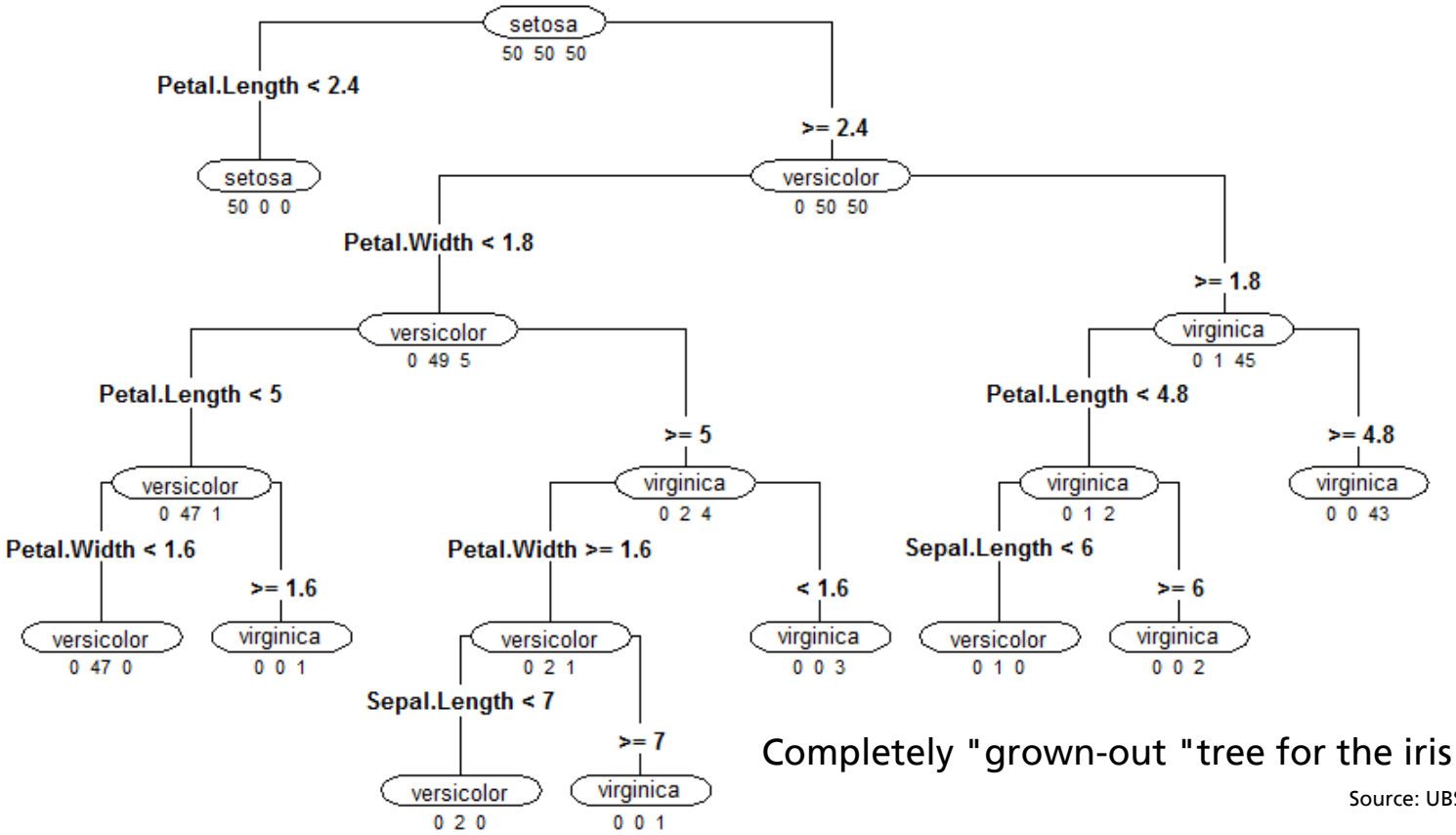
$$Impurity\ in\ long\ sepal\ subgroup = 1 - \left(\frac{22}{118}\right)^2 - \left(\frac{47}{118}\right)^2 - (\frac{49}{118})^2 = \mathbf{0.63}$$

$$Overall\ impurity\ after\ cut = \frac{32}{150} \times 0.22 + \frac{118}{150} \times 0.63 = \mathbf{0.55}$$

Lower impurity after the cut

# Over-fitting: The problem

- The rules we have seen so far will make the tree grow until every node is completely homogenous.
- This will typically lead to a big, complicated tree which over-fits the data.




Completely "grown-out "tree for the iris data

Source: UBS Quants

# Over-fitting: Overview of how R solves this

- Rpart has two stages to prevent over-fitting:

1. It does not allow the tree to grow out completely
2. It returns some statistics to show what a "sensible size" is.

# Over-fitting: Don't let the tree grow all the way out

Step 1

- To stop the tree from growing out completely, Rpart has some rules. The defaults cause the tree to stop when:
  - The node is very small (defaults to 20)
  - The best available split leads to a very small leaf (defaults to minsplit/3)
  - The best available split does not improve the fit by a certain amount (defaults to 0.01).
- You can change these defaults using the control variable in the rpart function, but it isn't typically necessary.
- This *can* still lead to an over-fitted tree, but it is much less likely.

# Over-fitting: Using the rpart output to choose the complexity of your tree (1)

## Step 2

- R will usually return an appropriately fitted tree, but the choice of how complex it should be is often more of an art than a science, so you may want a different tree to the default.
- The rpart function returns some statistics to help us:

```
myTree$cptable
```

```
      CP nsplit rel error xerror  xstd
1  0.500      0      1.00   1.21 0.0484
2  0.440      1      0.50   0.67 0.0609
3  0.020      2      0.06   0.10 0.0306
4  0.010      3      0.04   0.10 0.0306
5  0.005      6      0.01   0.07 0.0258
6 -1.000      8      0.00   0.07 0.0258
```

**Complexity parameter**

Lower values indicate a more complex tree.

**Relative error**

# of misclassified / # of misclassified in tree without any splits

**Cross validated relative error**

# of misclassified estimated from cross-validation / # of misclassified in tree without any splits

**Cross validated standard error**

Source: UBS Quants

UBS

# Over-fitting: Using the rpart output to choose the complexity of your tree (2)

| | CP | nsplit | rel error | xerror | xstd |
|---|---|---|---|---|---|
| 1 | 0.500 | 0 | 1.00 | 1.21 | 0.0484 |
| 2 | 0.440 | 1 | 0.50 | 0.67 | 0.0609 |
| 3 | 0.020 | 2 | 0.06 | 0.10 | 0.0306 |
| 4 | 0.010 | 3 | 0.04 | 0.10 | 0.0306 |
| **5** | **0.005** | **6** | **0.01** | **0.07** | **0.0258** |
| 6 | -1.000 | 8 | 0.00 | 0.07 | 0.0258 |

Source: UBS Quants

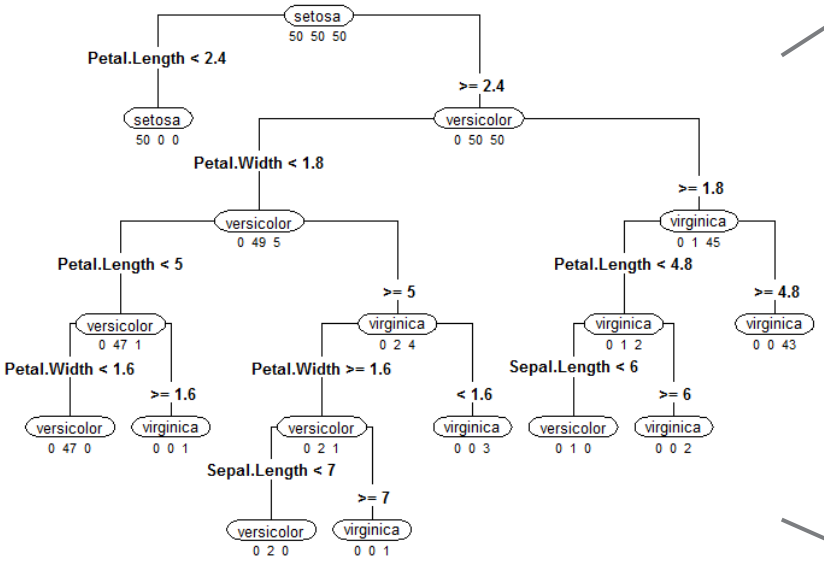
- We could choose the tree with the smallest cross-validated relative error i.e. the tree with six splits here.
- An alternative, more conservative approach is to take the smallest tree within one standard error of the smallest cross-validated relative error i.e. the smallest tree with:

$$\text{xerror} < 0.07 + 0.0258 = 0.0958$$

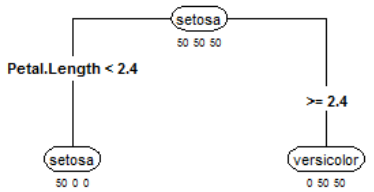
- … which would give us the tree with two splits
- Then we prune the tree to this size by specifying the cp parameter.

```
newTree = prune(myTree, cp=0.02)
```

# Over-fitting: Using the rpart output to choose the complexity of your tree (3)

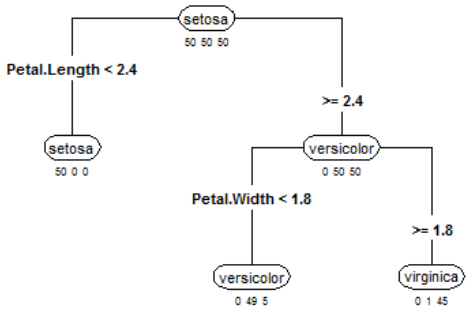



Prune at cp = 0.44



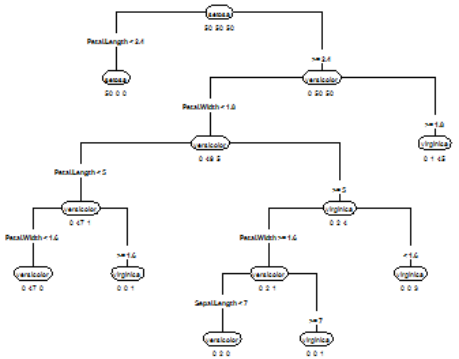

Too simple

Prune at cp = 0.02




**Just right**

Prune at cp = 0.005



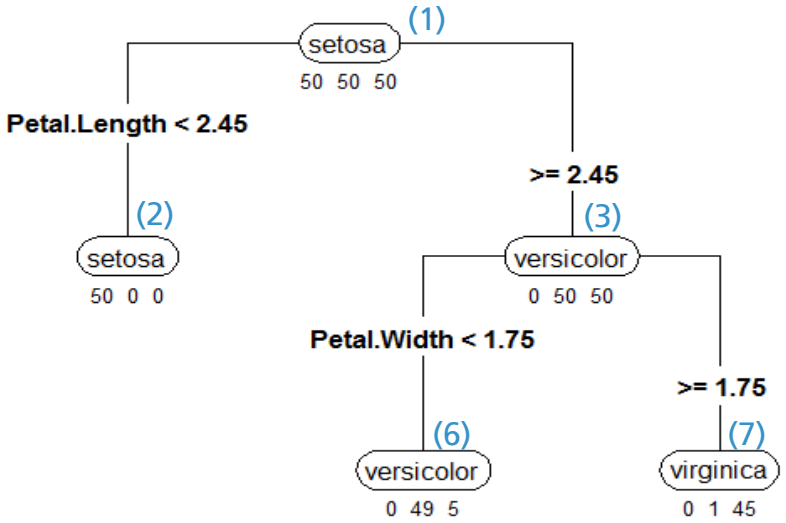
Too complex

Source: UBS Quants

# Trees in R: Create a decision tree

- Here is the R code to create a classification decision tree and the main output:

```
library(rpart)
myTree = rpart(Species ~., data=iris)
```




Source: UBS Quants

```
1) root 150 100 setosa (0.33 0.33 0.33)
  2) Petal.Length< 2.45 50   0 setosa (1.00 0.00 0.00) *
  3) Petal.Length>=2.45 100  50 versicolor (0.00 0.50 0.50)
    6) Petal.Width< 1.75 54   5 versicolor (0.00 0.91 0.09) *
    7) Petal.Width>=1.75 46   1 virginica (0.00 0.02 0.98) *
```

Node

Split

# of the data-points which go through that node

# of the data-points which are *incorrectly* predicted

Predicted classification

Breakdown of actual classifications at that node

Is this a leaf?

UBS

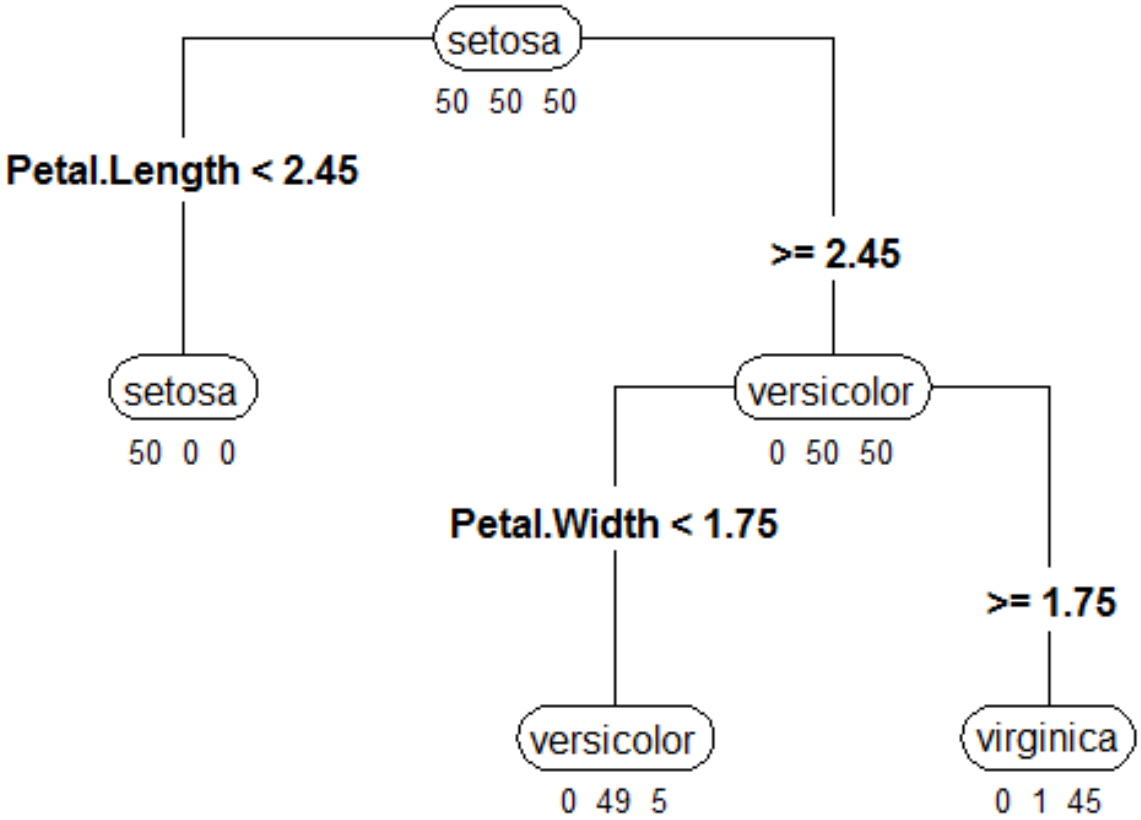
# Trees in R: Plot the tree

Simple plot

```
plot(myTree)
text(myTree)
```

Pretty plot

```
library(rpart.plot)
prp(myTree, type=4, extra=1, varlen=0, under=T, branch=1)
```




Source: UBS Quants

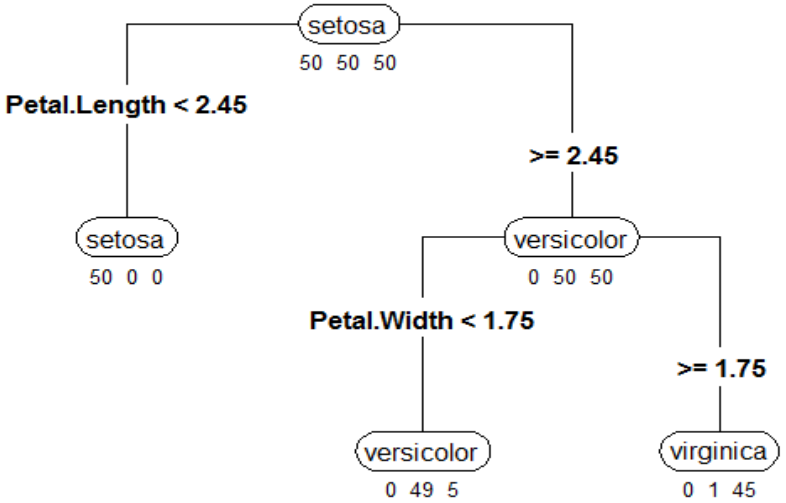
# Trees in R: Use the output

- To use your tree for predicting, put your out of sample data into the same format as your training data and use the prediction function:

```
x = data.frame(Sepal.Length=5, Sepal.Width=3.8, Petal.Length=2.5, Petal.Width=2.5)
predict(myTree, x)
```

```
  setosa versicolor virginica
1      0      0.022     0.978
```

- You get the vector of probabilities for each classification from the node of the tree that that data-point falls into.




Source: UBS Quants

Section 2

# Machine Learning Theory

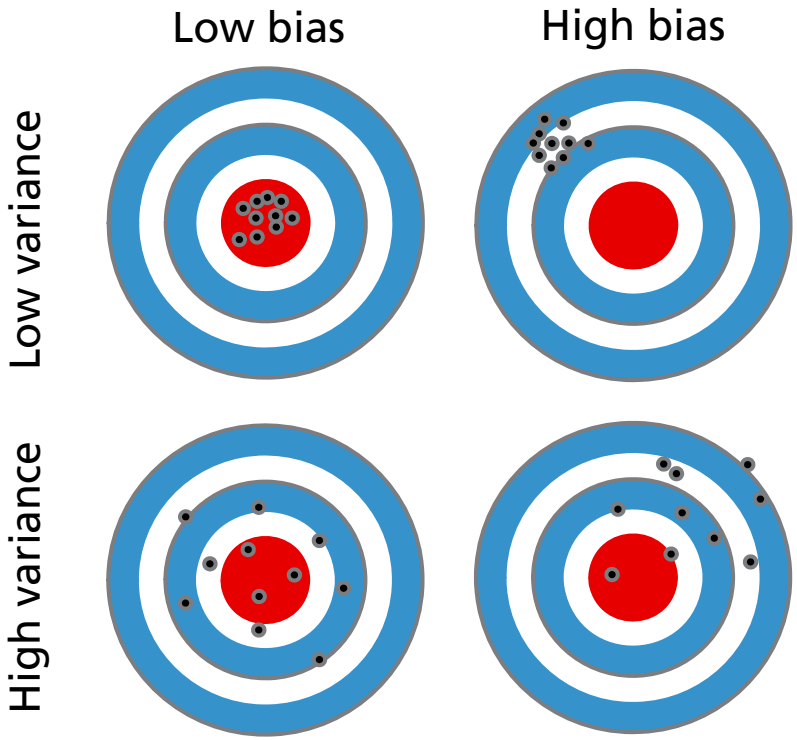
# Bias vs Variance (1)

- The **bias** is the difference between the expected prediction of the model and the correct value we are trying to predict.

  e.g. if you predict election outcomes from a survey of homeowners, the model will have high bias.

- The **variance** is how much the predictions for a given point vary between different realisations of the model.

  e.g. if you predict election outcomes from a survey of 20 people the model will have high variance.




Source: UBS Quants

# Bias vs Variance (2)

- There is typically a trade-off between bias and variance as you increase the model's complexity.

  e.g. suppose you fit a polynomial relationship between x and y with 100 data-points. As you increase the degree of the polynomial, you will decrease the bias, but increase the variance.

- If you over-fit a model you can expect to see low bias, but high variance.
- If you under-fit a model you can expect to see high bias but low variance.

- The total error on a model can be expressed as the sum of the variance and the bias$^2$, so finding the best model amounts to finding the sweet spot between variance and bias.
- Analysts often obsess over reducing bias, but variance is an equally important a source of error.

# Ensemble learning (1)

- Ensemble learning means combining a large number of simple models (often called "weak learners") to create an overall model. Typically we combine the simple models using the average or majority forecast from each model.
- Provided each simple model has better accuracy than random guessing and the models are diverse, then combining models should improve performance.
- Two of the most popular approaches are:

| **Bagging**<br>***B***ootstrap ***Agg***regat***ing*** | **Boosting** |
|---|---|
| Each model is created **independently**. This makes bagging parallelisable. | The models are created **iteratively**, so that a model focuses on forecasting the data instances which were misclassified by previous models. |
| Each model's forecast is treated equally. | Models' forecasts are weighted by their accuracy. |
| Works by reducing variance. So bagging is particularly useful for combining high variance, low bias models, such as over-fitted trees. | Works by reducing both variance and bias. So boosting is commonly used with high bias models e.g. 1-level trees ("stumps") |
| Example = Random forests | Example = adaBoost |

# Ensemble learning (2)

- Dietterich describes three conceptual reasons why ensembles can improve the performance of weak learners in his paper "Ensemble Methods in Machine Learning":
- Statistical

  When you have limited data, there are many different models which give similar accuracy on the training set. Using an ensemble allows you to reduce the risk of choosing the wrong model.
- Computational

  Many learning algorithms work by a local search, which can lead to them getting stuck in local optima. An ensemble approach allows you to try out multiple different start points which reduces this risk.
- Representational

  Sometimes by combining models you can produce hypotheses which are not available from individual models

  e.g. a decision tree will struggle to classify a diagonal boundary. The tree would have to be enormous to approximate it with any degree of accuracy. However, even a fairly small combination of trees can get a good approximation quickly.

Section 3

# Random Forests

# Overview of random forests (1)

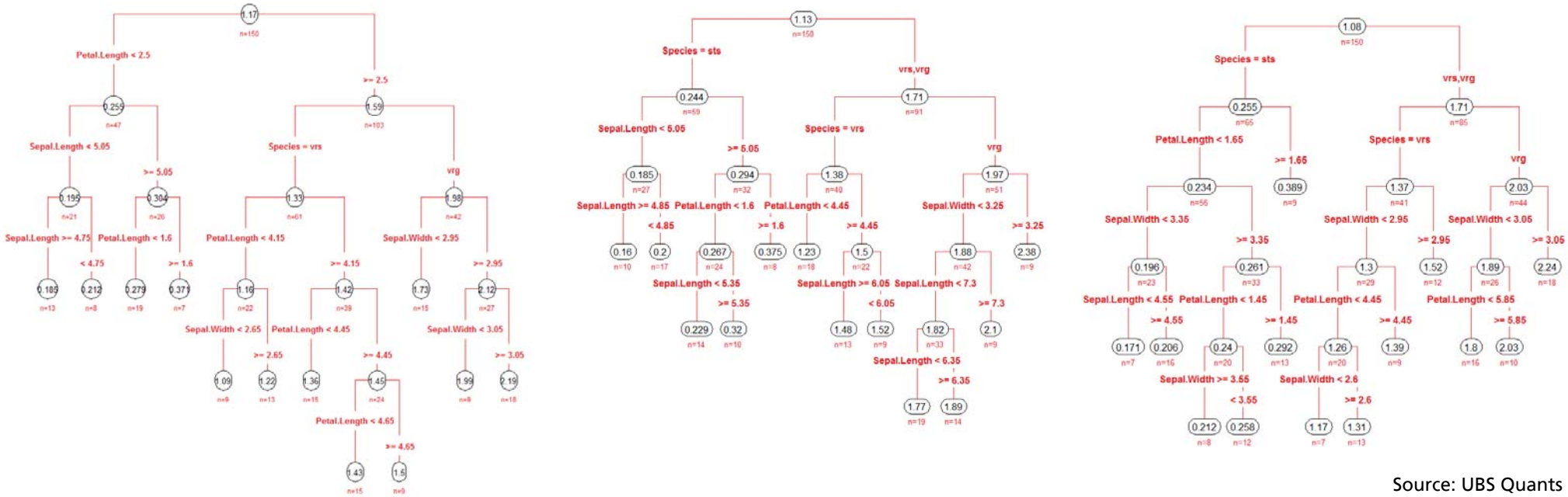
- Random forests (Breiman, 1999) is a bagging algorithm which uses decision trees.
- In brief, a random forest is a large collection of trees, with two additional sources of randomness.
- For each tree in the forest:
  1. Start with a bootstrap sample of the data.
  2. At each node, select M of the available features at random. Find the best feature and threshold to use at this node.
     - In R, the default value for M is set to √(# of variables).
  3. Continue this process with each of the two sub-groups until you have completely grown the tree out and each node is homogenous.

  Repeat until you have a large number of completely grown out trees (default in R is 500 but if there are a large number of variables in your dataset you may choose to increase that.)
- This gives us two sources of randomness; the bootstrap sample and the random selection of features at each node.

# Overview of random forests (2)

- To use the forest for prediction, you take the prediction from each of the individual trees and combine them with a majority vote.

Tree 1, Tree 2, …., Tree 500




Source: UBS Quants

# Why should a forest be better than a tree?

Intuitive argument:

- If you have multiple (somewhat independent) models with different pieces of useful information, then choosing just one of them, even if it is the best, means throwing away some of that information.
- If you instead combine models, you can keep the insights from all the models, which should produce a stronger, more robust overall model.

Bias vs Variance argument:

- In a random forest, the bias of the forest is equal to the bias of an individual (overgrown) tree, but the variance of the forest is much smaller than the variance of an individual tree, so the overall error is much lower.
- We ensure that each individual tree has low correlation to its peers by allowing them to grow all the way out, using bootstrap sampling and using random feature selection, so that the variance decreases more rapidly.

# Importance

- A random forest can also tell us which of our features is most important:
  - For each feature, we look across all the trees and find every node which uses that feature.
  - Then we compute how much each node which uses that feature decreases the Gini impurity.
  - We take the average.
- A higher value indicates the feature is very effective at distinguishing between classes and a lower value indicates the feature is less effective.
- In R, there is a function to do this automatically: `importance(myForest)`

| | MeanDecreaseGini |
|---|---|
| Sepal.Length | 9.9726 |
| Sepal.Width | 2.3978 |
| Petal.Length | 43.7739 |
| Petal.Width | 43.1792 |

Source: UBS Quants

- It can be helpful to use random forests to order a list of features to understand your data better, even if you then go on to use a different model.
- Be careful though; this statistic does not take into account the correlation between features. Two very highly correlated features will have similar importance, but there is little to be gained from including both in your model.

# Pros and cons of the random forests

## Pros

- Much more effective than an single decision tree.
- Very little pre-processing of your signals is required. Random forests can handle outliers, non-normal inputs and both categorical and continuous inputs.
- Easy to parallelise which allows for faster runtimes.
- Identifies important signals, which is helpful when you build other models.

## Cons

- Random forests are (to some extent) a black box. They are far harder to interpret and communicate than a simple decision tree.
- Can still struggle with severe class imbalances
- Random forests can struggle to generalise to new data if it lies outside the range of data it was trained on
  - e.g. if you are modelling the effect of inflation on equity returns and there is no deflationary period in your data, the model will probably not perform well at forecasting for deflationary periods.
- Random forests tend to be biased towards continuous variables or categorical variables with lots of splits.

# Random Forests: forecasting dividend cuts (1)

The problem:

- Forecast which stocks in the Russell 1000 will cut their dividend by more than 5% in the next 12 months
  - 5% requirement reduces "noise"
  - We only use ordinary dividends per share (no special dividends or stock dividends etc.)
  - We exclude firms which have had large corporate actions in the 12m before the ex-div date.

The features:

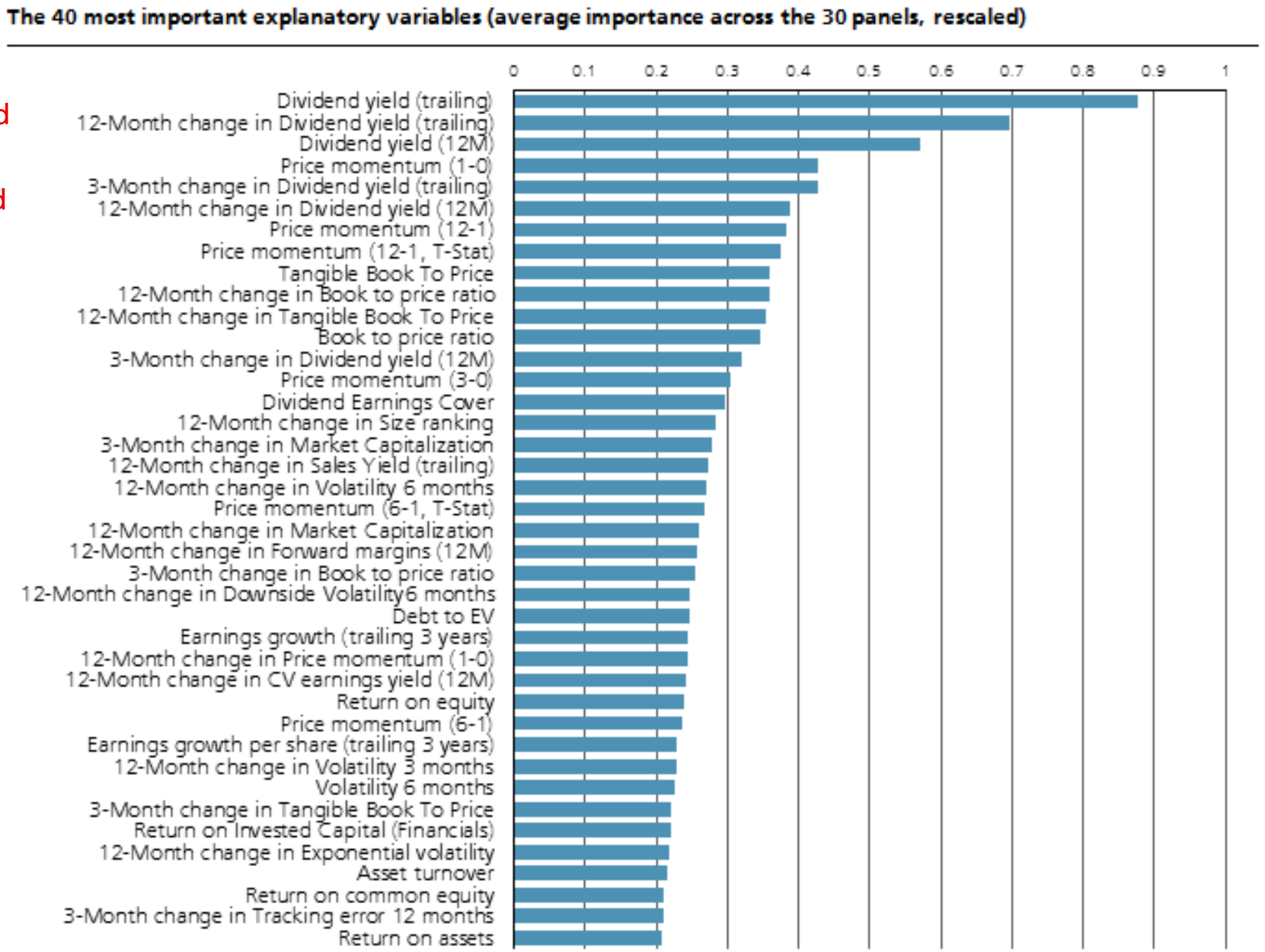
- We selected 116 factors from our daily factor database and supplemented each variable with its 3-month and 12-month changes. In total this gives us 348 explanatory variables
  - The model does not include information about the companies' dividend policies nor does it attempt to forecast changes in dividend policies
- We use 7 year rolling windows for the analysis
  - There are secular changes in dividend yields, so an expanding window is misleading
  - For long periods of time, almost no stocks cut their dividends so you need a long window.

The model:

- Random Forest with 2000 decision trees.

# Random Forests: forecasting dividend cuts (2)

1. Dividend yield
2. Changes in dividend yield
3. Price momentum
4. Book to price
5. Dividend cover
6. Debt to Enterprise Value
7. Historic earnings growth
8. Profitability
9. Volatility

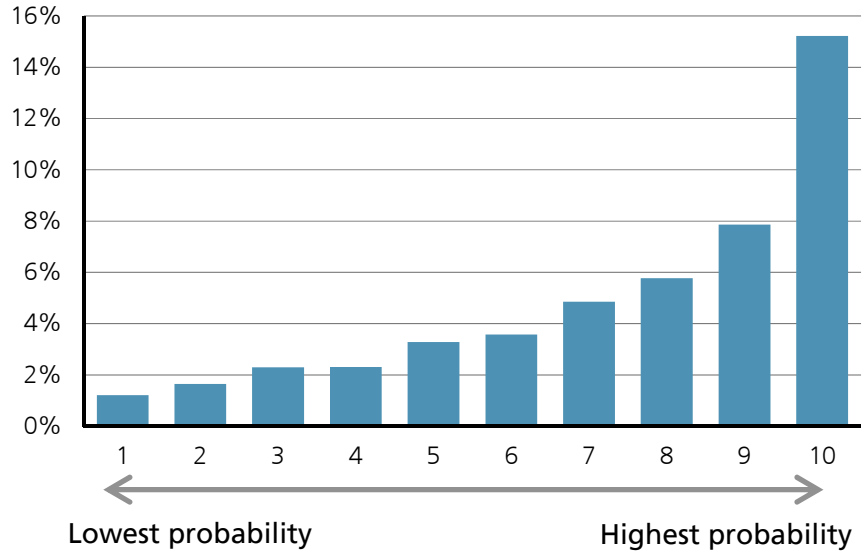



Source: UBS Quantitative Research

UBS

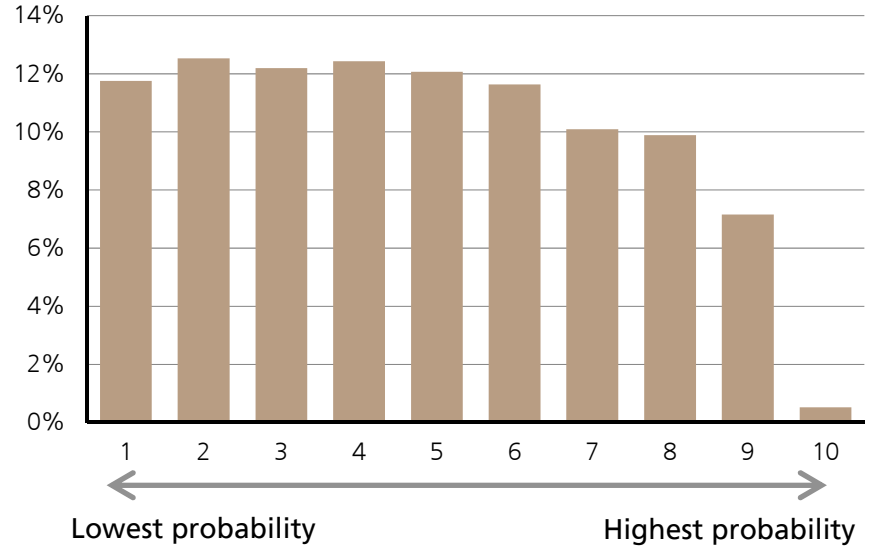
# Random Forests: forecasting dividend cuts (3)

**Proportion of dividend reduction in each decile**




Source: UBS Quantitative Research. US largest 1000 stocks.

**Average DPS growth rate in each decile**




Source: UBS Quantitative Research. Trimmed average. US largest 1000 stocks.

- More than 15% of the companies with the highest estimated likelihood of a dividend cut ended up reducing their dividends compared to just 4.8% across all the deciles.
- The companies with the highest expected probability of a dividend cut grew their dividend per share by an average of less than 2.8% over the following 12 months compared to an average growth rate of more than 10% across all the companies.

# Random forests in R: Create a forest

- Here is the R code to create a simple classification random forest and the main output:

```
library(randomForest)
myForest = randomForest(Species ~., data=iris)
```

```
Call:
 randomForest(formula = Species ~ ., data = iris)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 5.33%
Confusion matrix:
           setosa versicolor virginica class.error
setosa         50          0         0        0.00
versicolor      0         47         3        0.06
virginica       0          5        45        0.10
```

**Out-of-bag estimate of error rate**

For each entry $X_i$ in the dataset, predict that $y_i$ using only the trees which do not include $(X_i, y_i)$. Find the error.

**Confusion matrix**

This shows how the training data would be classified / misclassified by the out-of-bag estimation.

Source: UBS Quants

# Random forests in R: Use the output

- To use your random forest for predicting then, just as before, put your out of sample data into the same format as your training data and use the predict function:

```
x = data.frame(Sepal.Length=5, Sepal.Width=3.8, Petal.Length=2.5, Petal.Width=2.5)
predict(myForest, x, "prob")
```

```
  setosa versicolor virginica
1   0.21      0.392     0.398
```

Source: UBS Quants

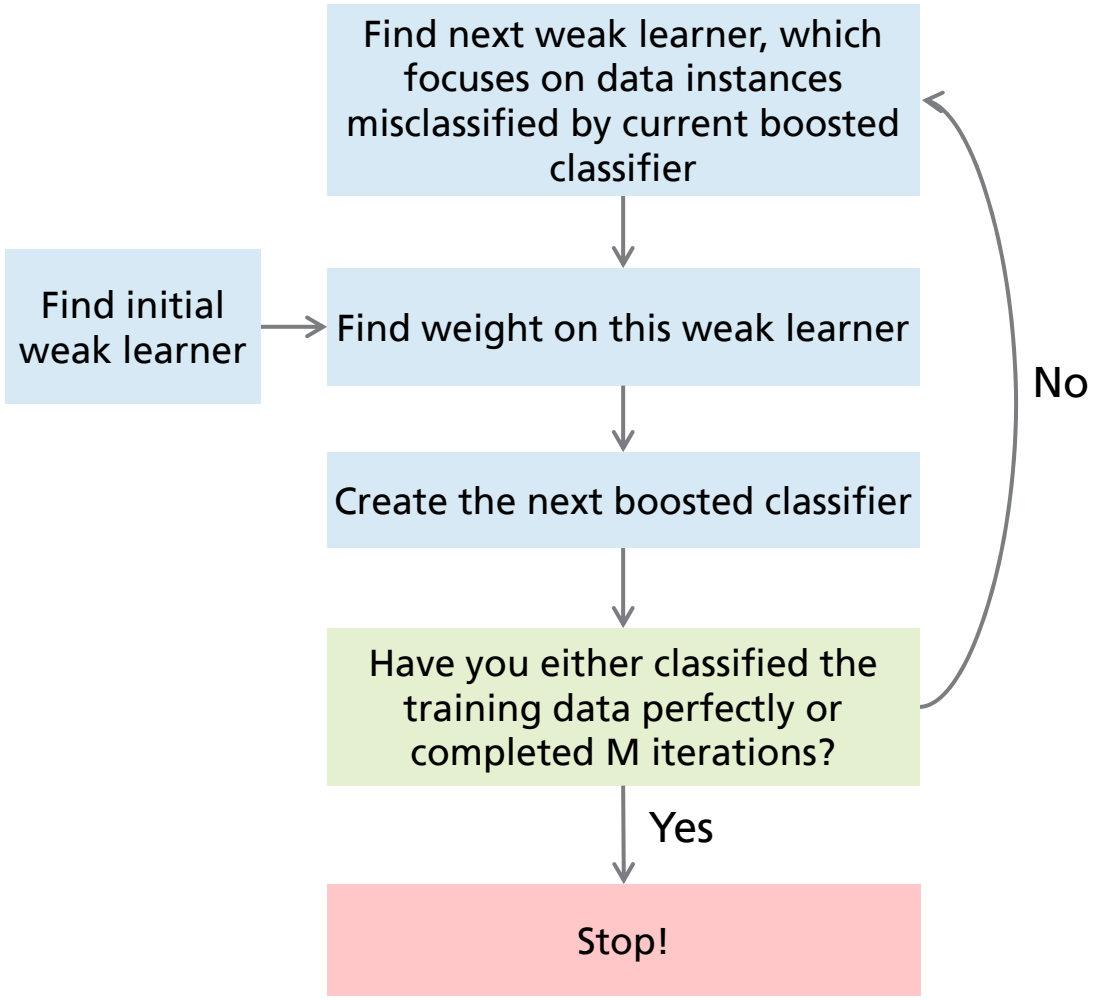
- Once again, you get the vector of probabilities for each classification.

Section 4

# AdaBoost

# AdaBoost: algorithm in words

- AdaBoost is a boosting algorithm which often (but not always) uses decision trees.
- There are many variants of AdaBoost, but we will focus on the basic, most common version.
- We *iteratively* create a series of weak classifiers (often decision trees which are constrained to have only one node) on the dataset.
- As we go, we create a boosted classifier made up of a *linear combination* of all the weak classifiers so far.
- Each new weak model is built to be expert at instances handled incorrectly by the current boosted classifier.




Source: UBS Quants

# AdaBoost: algorithm in maths

- Some definitions:
  - $w_i$ is the weight of data instance i.
  - $\kappa_m$ is the m$^{th}$ weak classifier. This is often a heavily pruned decision tree.
  - $\alpha_m$ is the weight of the m$^{th}$ weak classifier in the boosted classifier
  - $C_m$ is the m$^{th}$ boosted classifier. This is defined as a linear combination of the weak classifiers so far:

$$C_m(x_i) = \alpha_1\kappa_1(x_i) + \cdots + \alpha_m\kappa_m(x_i)$$

Start with equal weights

Initialise the weights, so that $w_i = 1$ for all i

For m in 1 to M:

Each new weak classifier focuses on data instances with bigger weights - those currently being misclassified

Find the weak classifier, $\kappa_m$, which minimises the weighted error:

$$error = \sum_{i:\, \kappa_m(x_i) \neq y_i} w_i \Big/ \sum_{i=1}^{N} w_i$$

Weak classifiers with smaller weighted errors will have a bigger impact on the boosted classifier

Define $\alpha_m = \frac{1}{2}\ln(\frac{1-error}{error})$

Define $C_m$, the m$^{th}$ boosted classifier as $C_m(x_i) = \alpha_1\kappa_1(x_i) + \cdots + \alpha_m\kappa_m(x_i)$

Give greater weight to data instances which are misclassified by the boosted classifier

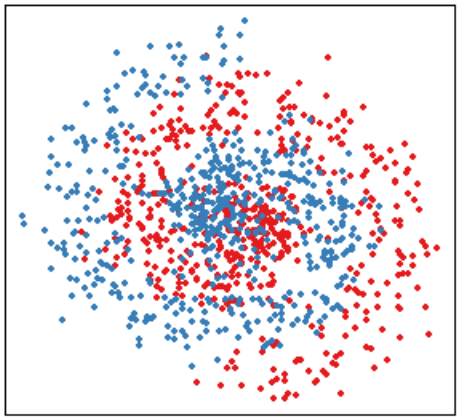
Update the weights so that $w_i = \exp(-y_i C_m(x_i))$ for all i

End

Source: UBS Quants

# AdaBoost: algorithm in pictures

Start with equal weights

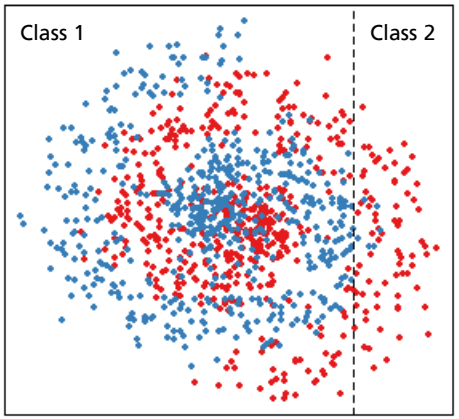


Find the first weak classifier, $\kappa_1$

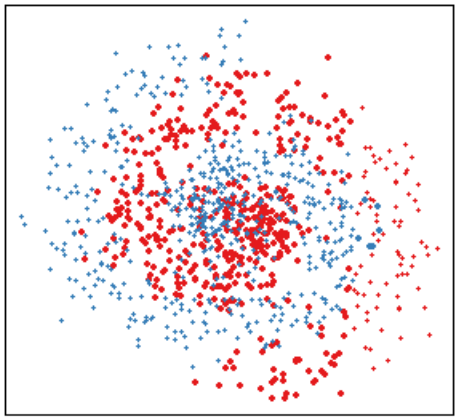
Weighted error = 43%

$\alpha_1 = 0.13$

$C_1 = 0.13\kappa_1$




Update the weights so that misclassified data instances get greater weights.



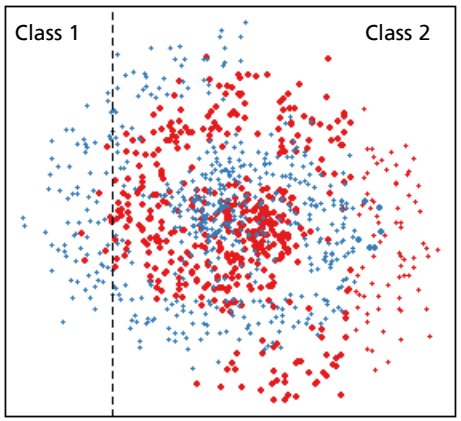
Find the second weak classifier, $\kappa_2$

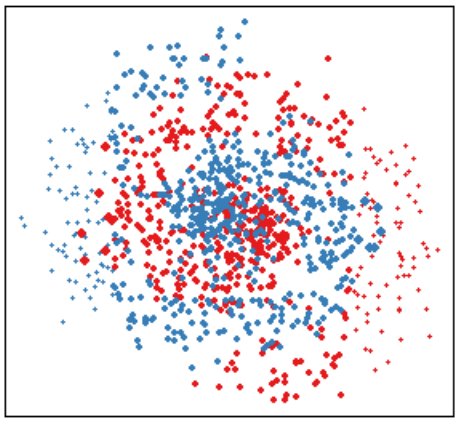
Weighted error = 39%

$\alpha_2 = 0.22$

$C_s = 0.13\kappa_1 + 0.224\kappa_2$

Update the weights so that misclassified data instances get greater weights.



...

UBS

Source: UBS Quants

# Pros and cons of AdaBoost

## Pros

- Very robust to data inputs – doesn't need to be normal or continuous
- Unlikely to over-fit
- Can reduce both bias and variance

## Cons

- Very sensitive to incorrect classification of data in its training set.
- Sometimes struggles with very noisy data.
- Cannot be run in parallel.

# AdaBoost in R: Create your model

- Here is the R code to create an adaBoost model. Currently the package ada only supports classification problems with two classes.

```
library(ada)
myAdaBoost = ada(Species~., data=subset(iris, Species!='setosa'))
```

```
Call:
ada(Species ~ ., data = subset(iris, Species != "setosa"))
Loss: exponential Method: discrete   Iteration: 50


Final Confusion Matrix for Data:
            Final Prediction
True value   versicolor virginica
  versicolor         47         3
  virginica           2        48


Train Error: 0.05
Out-Of-Bag Error:  0.04  iteration= 11
```

Confusion matrix

This shows how the training data would be classified / misclassified.

Error estimates

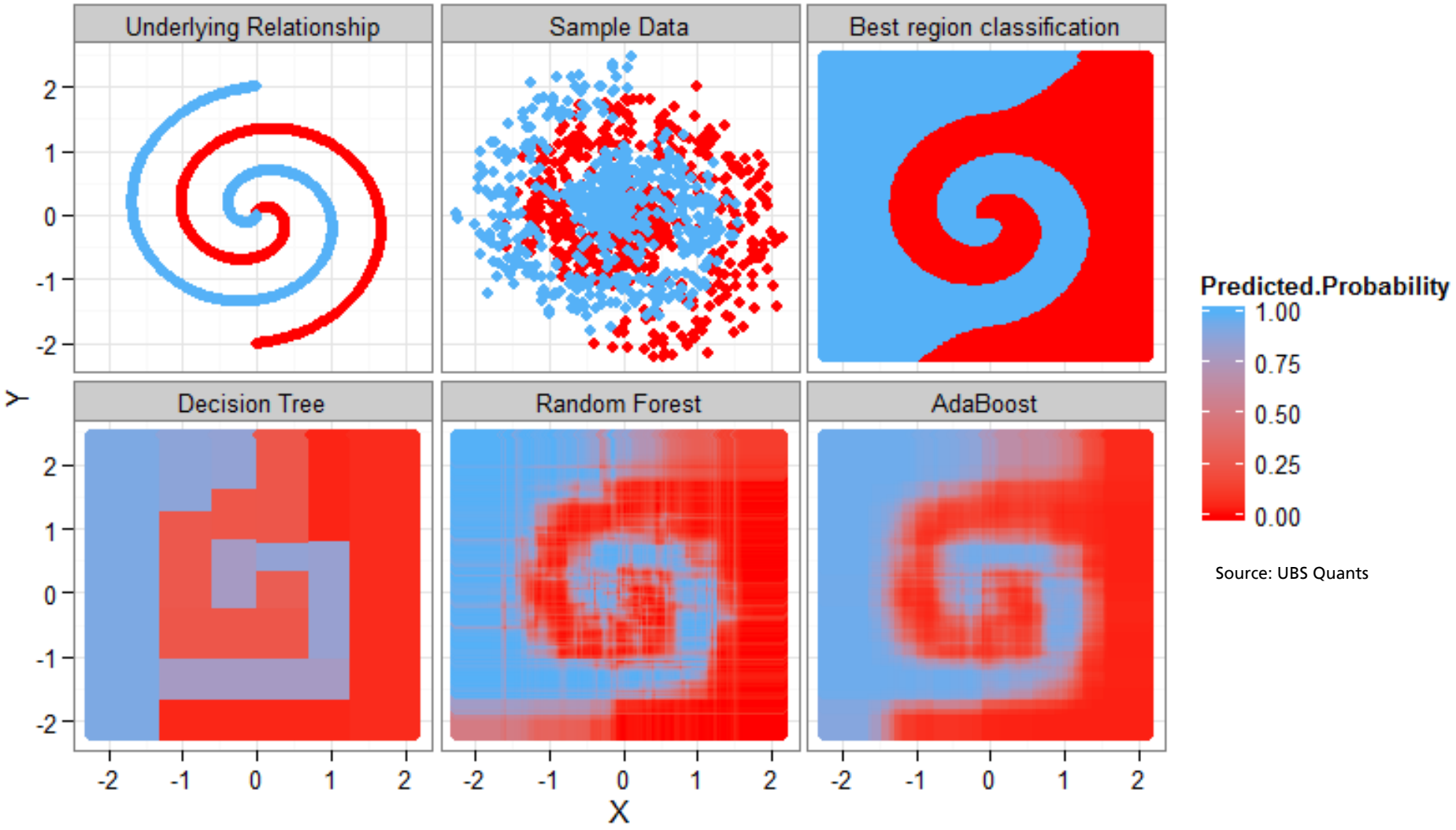
Source: UBS Quants

Section 5

# Machine Learning Horse Race

UBS

# Machine learning horse race




Source: UBS Quants

| | Decision Tree | Random forest | AdaBoost |
|---|---|---|---|
| Out of sample accuracy | 70% | 76% | 76% |
| Time taken | 0.05 secs | 0.435 secs | 0.863 secs |

Source: UBS Quants

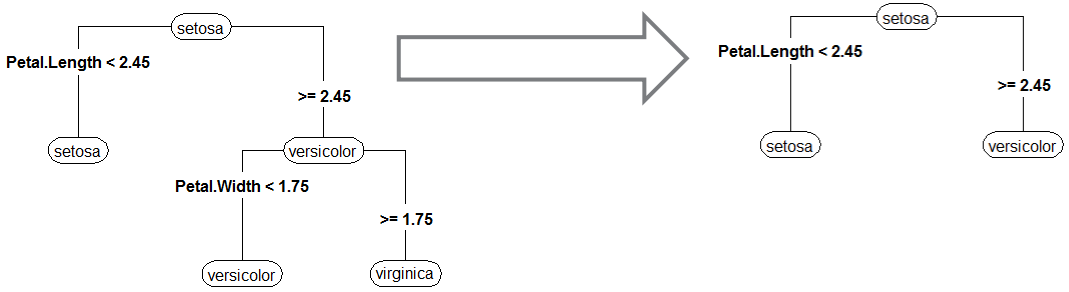
Section 6

# Appendix

# Over-fitting: Optional extra details (1)

First some definitions:

- A sub-tree of a tree T is any smaller tree you can make by pruning one or more nodes, for example:




- The big, over-grown tree which is step 1 in R's process is called $T_{max.}$
- We define: $\mathbf{R_\alpha(T)} = \%\ \text{of errors} + \alpha \times \#\ \text{of leaves}$

  And: $\mathbf{T_\alpha} = \text{subtree of } T_{Max} \text{ which minimises } R_\alpha(T_\alpha)$
- As you increase alpha, so you penalise bigger trees more, you get a series of nested sub-trees of decreasing complexity. This series is finite and one of these trees will be our optimal tree.
- $T_0$ is simply $T_{max}$, the fully grown tree, and for large enough $\alpha$, $T_\alpha$ will be the minimal tree with no splits (just the root).

# Over-fitting: Optional extra details (2)

- We use cross-validation to find the best alpha to use and hence the best sub-tree which has both a good fit and is parsimonious.

- First: Divide the data into 10 sub-sets $X_1$ … $X_{10}$.

  For i in 1 to 10

  > Our data-set is all the data except $X_i$ i.e. nine-tenths of the full data-set.
  >
  > Find $T_{max}$, the (overgrown) tree for this data-set.
  >
  > For any alpha
  >
  > > Find $T_\alpha$, the sub-tree of $T_{max}$ which minimises the error adjusted for complexity
  > >
  > > $$R_\alpha(T_\alpha) = \%\ \text{of errors} + \ \alpha \times \#\ \text{of leaves}$$
  > >
  > > Then, using the out-of-sample data $X_i$ find $R_\alpha(T_\alpha)$
  >
  > End

  End

- Then: For any alpha

  > Find the average out-of-sample value of the error adjusted for complexity, $R_\alpha(T_\alpha)$.

  End

  The "best" alpha is the one with the smallest average.

- Finally: Use all of the data, X, to find $T_{max}$ (the overgrown tree).

  Then, using the best alpha, find $T_\alpha$, the sub-tree of $T_{max}$ which minimises $R_\alpha(T_\alpha)$

# Analyst Certification

## Analyst Certification

Each research analyst primarily responsible for the content of this research report, in whole or in part, certifies that with respect to each security or issuer that the analyst covered in this report: (1) all of the views expressed accurately reflect his or her personal views about those securities or issuers and were prepared in an independent manner, including with respect to UBS, and (2) no part of his or her compensation was, is, or will be, directly or indirectly, related to the specific recommendations or views expressed by that research analyst in the research report.

# Required Disclosures

This report has been prepared by UBS Limited, an affiliate of UBS AG. UBS AG, its subsidiaries, branches and affiliates are referred to herein as UBS.

For information on the ways in which UBS manages conflicts and maintains independence of its research product; historical performance information; and certain additional disclosures concerning UBS research recommendations, please visit www.ubs.com/disclosures. The figures contained in performance charts refer to the past; past performance is not a reliable indicator of future results. Additional information will be made available upon request. UBS Securities Co. Limited is licensed to conduct securities investment consultancy businesses by the China Securities Regulatory Commission. UBS acts or may act as principal in the debt securities (or in related derivatives) that may be the subject of this report.

**UBS Investment Research: Global Equity Rating Definitions**

| 12-Month Rating | Definition | Coverage[1] | IB Services[2] |
|---|---|---|---|
| **Buy** | FSR is > 6% above the MRA. | 48% | 36% |
| **Neutral** | FSR is between -6% and 6% of the MRA. | 39% | 28% |
| **Sell** | FSR is > 6% below the MRA. | 12% | 22% |
| **Short-Term Rating** | **Definition** | **Coverage[3]** | **IB Services[4]** |
| **Buy** | Stock price expected to rise within three months from the time the rating was assigned because of a specific catalyst or event. | <1% | <1% |
| **Sell** | Stock price expected to fall within three months from the time the rating was assigned because of a specific catalyst or event. | <1% | <1% |

Source: UBS. Rating allocations are as of 31 December 2015.
1:Percentage of companies under coverage globally within the 12-month rating category.
2:Percentage of companies within the 12-month rating category for which investment banking (IB) services were provided within the past 12 months.
3:Percentage of companies under coverage globally within the Short-Term rating category.
4:Percentage of companies within the Short-Term rating category for which investment banking (IB) services were provided within the past 12 months.

**KEY DEFINITIONS:** **Forecast Stock Return (FSR)** is defined as expected percentage price appreciation plus gross dividend yield over the next 12 months. **Market Return Assumption (MRA)** is defined as the one-year local market interest rate plus 5% (a proxy for, and not a forecast of, the equity risk premium). **Under Review (UR)** Stocks may be flagged as UR by the analyst, indicating that the stock's price target and/or rating are subject to possible change in the near term, usually in response to an event that may affect the investment case or valuation. **Short-Term Ratings** reflect the expected near-term (up to three months) performance of the stock and do not reflect any change in the fundamental view or investment case. **Equity Price Targets** have an investment horizon of 12 months.

**EXCEPTIONS AND SPECIAL CASES: UK and European Investment Fund ratings and definitions are: Buy:** Positive on factors such as structure, management, performance record, discount; **Neutral:** Neutral on factors such as structure, management, performance record, discount; **Sell:** Negative on factors such as structure, management, performance record, discount. **Core Banding Exceptions (CBE):** Exceptions to the standard +/-6% bands may be granted by the Investment Review Committee (IRC). Factors considered by the IRC include the stock's volatility and the credit spread of the respective company's debt. As a result, stocks deemed to be very high or low risk may be subject to higher or lower bands as they relate to the rating. When such exceptions apply, they will be identified in the Company Disclosures table in the relevant research piece.

Research analysts contributing to this report who are employed by any non-US affiliate of UBS Securities LLC are not registered/qualified as research analysts with FINRA. Such analysts may not be associated persons of UBS Securities LLC and therefore are not subject to the FINRA restrictions on communications with a subject company, public appearances, and trading securities held by a research analyst account. The name of each affiliate and analyst employed by that affiliate contributing to this report, if any, follows.

**UBS Limited:** Claire Jones, Nick Baltas

Unless otherwise indicated, please refer to the Valuation and Risk sections within the body of this report.

# Compendium Disclosure

For a complete set of disclosure statements associated with the companies discussed in this report, including information on valuation and risk, please contact:

UBS Securities LLC,
1285 Avenue of Americas,
New York, NY 10019, USA.

Attention: Publishing Administration

# Global Disclaimer

This document has been prepared by UBS Limited, an affiliate of UBS AG. UBS AG, its subsidiaries, branches and affiliates are referred to herein as UBS.

Global Research is provided to our clients through UBS Neo, the UBS Client Portal and UBS.com (each a "System"). It may also be made available through third party vendors and distributed by UBS and/or third parties via e-mail or alternative electronic means. The level and types of services provided by Global Research to a client may vary depending upon various factors such as a client's individual preferences as to the frequency and manner of receiving communications, a client's risk profile and investment focus and perspective (e.g. market wide, sector specific, long-term, short-term, etc.), the size and scope of the overall client relationship with UBS and legal and regulatory constraints.

When you receive Global Research through a System, your access and/or use of such Global Research is subject to this Global Research Disclaimer and to the terms of use governing the applicable System.

When you receive Global Research via a third party vendor, e-mail or other electronic means, your use shall be subject to this Global Research Disclaimer and to UBS's Terms of Use/Disclaimer (http://www.ubs.com/global/en/legalinfo2/disclaimer.html). By accessing and/or using Global Research in this manner, you are indicating that you have read and agree to be bound by our Terms of Use/Disclaimer. In addition, you consent to UBS processing your personal data and using cookies in accordance with our Privacy Statement (http://www.ubs.com/global/en/legalinfo2/privacy.html) and cookie notice (http://www.ubs.com/global/en/homepage/cookies/cookie-management.html).

**If you receive Global Research, whether through a System or by any other means, you agree that you shall not copy, revise, amend, create a derivative work, transfer to any third party, or in any way commercially exploit any UBS research provided via Global Research or otherwise, and that you shall not extract data from any research or estimates provided to you via Global Research or otherwise, without the prior written consent of UBS.**

For access to all available Global Research on UBS Neo and the Client Portal, please contact your UBS sales representative.

This document is for distribution only as may be permitted by law. It is not directed to, or intended for distribution to or use by, any person or entity who is a citizen or resident of or located in any locality, state, country or other jurisdiction where such distribution, publication, availability or use would be contrary to law or regulation or would subject UBS to any registration or licensing requirement within such jurisdiction. It is published solely for information purposes; it is not an advertisement nor is it a solicitation or an offer to buy or sell any financial instruments or to participate in any particular trading strategy. No representation or warranty, either expressed or implied, is provided in relation to the accuracy, completeness or reliability of the information contained in this document ('the Information'), except with respect to Information concerning UBS. The Information is not intended to be a complete statement or summary of the securities, markets or developments referred to in the document. UBS does not undertake to update or keep current the Information. Any opinions expressed in this document may change without notice and may differ or be contrary to opinions expressed by other business areas or groups of UBS. Any statements contained in this report attributed to a third party represent UBS's interpretation of the data, information and/or opinions provided by that third party either publicly or through a subscription service, and such use and interpretation have not been reviewed by the third party.

Nothing in this document constitutes a representation that any investment strategy or recommendation is suitable or appropriate to an investor's individual circumstances or otherwise constitutes a personal recommendation. Investments involve risks, and investors should exercise prudence and their own judgement in making their investment decisions. The financial instruments described in the document may not be eligible for sale in all jurisdictions or to certain categories of investors. Options, derivative products and futures are not suitable for all investors, and trading in these instruments is considered risky. Mortgage and asset-backed securities may involve a high degree of risk and may be highly volatile in response to fluctuations in interest rates or other market conditions. Foreign currency rates of exchange may adversely affect the value, price or income of any security or related instrument referred to in the document. For investment advice, trade execution or other enquiries, clients should contact their local sales representative.

The value of any investment or income may go down as well as up, and investors may not get back the full (or any) amount invested. Past performance is not necessarily a guide to future performance. Neither UBS nor any of its directors, employees or agents accepts any liability for any loss (including investment loss) or damage arising out of the use of all or any of the Information.

Any prices stated in this document are for information purposes only and do not represent valuations for individual securities or other financial instruments. There is no representation that any transaction can or could have been effected at those prices, and any prices do not necessarily reflect UBS's internal books and records or theoretical model-based valuations and may be based on certain assumptions. Different assumptions by UBS or any other source may yield substantially different results.

This document and the Information are produced by UBS as part of its research function and are provided to you solely for general background information. UBS has no regard to the specific investment objectives, financial situation or particular needs of any specific recipient. In no circumstances may this document or any of the Information be used for any of the following purposes:

(i) valuation or accounting purposes;

(ii) to determine the amounts due or payable, the price or the value of any financial instrument or financial contract; or

(iii) to measure the performance of any financial instrument.

By receiving this document and the Information you will be deemed to represent and warrant to UBS that you will not use this document or any of the Information for any of the above purposes or otherwise rely upon this document or any of the Information.

Research will initiate, update and cease coverage solely at the discretion of UBS Investment Bank Research Management. The analysis contained in this document is based on numerous assumptions. Different assumptions could result in materially different results. The analyst(s) responsible for the preparation of this document may interact with trading desk personnel, sales personnel and other parties for the purpose of gathering, applying and interpreting market information. UBS relies on information barriers to control the flow of information contained in one or more areas within UBS into other areas, units, groups or affiliates of UBS. The compensation of the analyst who prepared this document is determined exclusively by research management and senior management (not including investment banking). Analyst compensation is not based on investment banking revenues; however, compensation may relate to the revenues of UBS Investment Bank as a whole, of which investment banking, sales and trading are a part.

For financial instruments admitted to trading on an EU regulated market: UBS AG, its affiliates or subsidiaries (excluding UBS Securities LLC) acts as a market maker or liquidity provider (in accordance with the interpretation of these terms in the UK) in the financial instruments of the issuer save that where the activity of liquidity provider is carried out in accordance with the definition given to it by the laws and regulations of any other EU jurisdictions, such information is separately disclosed in this document. For financial instruments admitted to trading on a non-EU regulated market: UBS may act as a market maker save that where this activity is carried out in the US in accordance with the definition given to it by the relevant laws and regulations, such activity will be specifically disclosed in this document. UBS may have issued a warrant the value of which is based on one or more of the financial instruments referred to in the document. UBS and its affiliates and employees may have long or short positions, trade as principal and buy and sell in instruments or derivatives identified herein; such transactions or positions may be inconsistent with the opinions expressed in this document.

# Global Disclaimer (continued)

**United Kingdom and the rest of Europe:** Except as otherwise specified herein, this material is distributed by UBS Limited to persons who are eligible counterparties or professional clients. UBS Limited is authorised by the Prudential Regulation Authority and regulated by the Financial Conduct Authority and the Prudential Regulation Authority. **France:** Prepared by UBS Limited and distributed by UBS Limited and UBS Securities France S.A. UBS Securities France S.A. is regulated by the ACPR (Autorité de Contrôle Prudentiel et de Résolution) and the Autorité des Marchés Financiers (AMF). Where an analyst of UBS Securities France S.A. has contributed to this document, the document is also deemed to have been prepared by UBS Securities France S.A. **Germany:** Prepared by UBS Limited and distributed by UBS Limited and UBS Deutschland AG. UBS Deutschland AG is regulated by the Bundesanstalt fur Finanzdienstleistungsaufsicht (BaFin). **Spain:** Prepared by UBS Limited and distributed by UBS Limited and UBS Securities España SV, SA. UBS Securities España SV, SA is regulated by the Comisión Nacional del Mercado de Valores (CNMV). **Turkey:** Distributed by UBS Limited. No information in this document is provided for the purpose of offering, marketing and sale by any means of any capital market instruments and services in the Republic of Turkey. Therefore, this document may not be considered as an offer made or to be made to residents of the Republic of Turkey. UBS AG is not licensed by the Turkish Capital Market Board under the provisions of the Capital Market Law (Law No. 6362). Accordingly, neither this document nor any other offering material related to the instruments/services may be utilized in connection with providing any capital market services to persons within the Republic of Turkey without the prior approval of the Capital Market Board. However, according to article 15 (d) (ii) of the Decree No. 32, there is no restriction on the purchase or sale of the securities abroad by residents of the Republic of Turkey. **Poland:** Distributed by UBS Limited (spolka z ograniczona odpowiedzialnoscia) Oddzial w Polsce regulated by the Polish Financial Supervision Authority. Where an analyst of UBS Limited (spolka z ograniczona odpowiedzialnoscia) Oddzial w Polsce has contributed to this document, the document is also deemed to have been prepared by UBS Limited (spolka z ograniczona odpowiedzialnoscia) Oddzial w Polsce. **Russia:** Prepared and distributed by UBS Bank (OOO). **Switzerland:** Distributed by UBS AG to persons who are institutional investors only. UBS AG is regulated by the Swiss Financial Market Supervisory Authority (FINMA). **Italy:** Prepared by UBS Limited and distributed by UBS Limited and UBS Italia Sim S.p.A. UBS Italia Sim S.p.A. is regulated by the Bank of Italy and by the Commissione Nazionale per le Società e la Borsa (CONSOB). Where an analyst of UBS Italia Sim S.p.A. has contributed to this document, the document is also deemed to have been prepared by UBS Italia Sim S.p.A. **South Africa:** Distributed by UBS South Africa (Pty) Limited (Registration No. 1995/011140/07), an authorised user of the JSE and an authorised Financial Services Provider (FSP 7328). **Israel:** This material is distributed by UBS Limited. UBS Limited is authorised by the Prudential Regulation Authority and regulated by the Financial Conduct Authority and the Prudential Regulation Authority. UBS Securities Israel Ltd is a licensed Investment Marketer that is supervised by the Israel Securities Authority (ISA). UBS Limited and its affiliates incorporated outside Israel are not licensed under the Israeli Advisory Law. UBS Limited is not covered by insurance as required from a licensee under the Israeli Advisory Law. UBS may engage among others in issuance of Financial Assets or in distribution of Financial Assets of other issuers for fees or other benefits. UBS Limited and its affiliates may prefer various Financial Assets to which they have or may have Affiliation (as such term is defined under the Israeli Advisory Law). Nothing in this Material should be considered as investment advice under the Israeli Advisory Law. This Material is being issued only to and/or is directed only at persons who are Eligible Clients within the meaning of the Israeli Advisory Law, and this material must not be relied on or acted upon by any other persons. **Saudi Arabia:** This document has been issued by UBS AG (and/or any of its subsidiaries, branches or affiliates), a public company limited by shares, incorporated in Switzerland with its registered offices at Aeschenvorstadt 1, CH-4051 Basel and Bahnhofstrasse 45, CH-8001 Zurich. This publication has been approved by UBS Saudi Arabia (a subsidiary of UBS AG), a Saudi closed joint stock company incorporated in the Kingdom of Saudi Arabia under commercial register number 1010257812 having its registered office at Tatweer Towers, P.O. Box 75724, Riyadh 11588, Kingdom of Saudi Arabia. UBS Saudi Arabia is authorized and regulated by the Capital Market Authority to conduct securities business under license number 08113-37. **Dubai:** The information distributed by UBS AG Dubai Branch is intended for Professional Clients only and is not for further distribution within the United Arab Emirates. **United States:** Distributed to US persons by either UBS Securities LLC or by UBS Financial Services Inc., subsidiaries of UBS AG; or by a group, subsidiary or affiliate of UBS AG that is not registered as a US broker-dealer (a **'non-US affiliate'** ) to major US institutional investors only. UBS Securities LLC or UBS Financial Services Inc. accepts responsibility for the content of a document prepared by another non-US affiliate when distributed to US persons by UBS Securities LLC or UBS Financial Services Inc. All transactions by a US person in the securities mentioned in this document must be effected through UBS Securities LLC or UBS Financial Services Inc., and not through a non-US affiliate. UBS Securities LLC is not acting as a municipal advisor to any municipal entity or obligated person within the meaning of Section 15B of the Securities Exchange Act (the "Municipal Advisor Rule"), and the opinions or views contained herein are not intended to be, and do not constitute, advice within the meaning of the Municipal Advisor Rule. **Canada:** Distributed by UBS Securities Canada Inc., a registered investment dealer in Canada and a Member-Canadian Investor Protection Fund, or by another affiliate of UBS AG that is registered to conduct business in Canada or is otherwise exempt from registration. **Brazil:** Except as otherwise specified herein, this material is prepared by UBS Brasil CCTVM S.A. to persons who are eligible investors residing in Brazil, which are considered to be: (i) financial institutions, (ii) insurance firms and investment capital companies, (iii) supplementary pension entities, (iv) entities that hold financial investments higher than R$300,000.00 and that confirm the status of qualified investors in written, (v) investment funds, (vi) securities portfolio managers and securities consultants duly authorized by Comissão de Valores Mobiliários (CVM), regarding their own investments, and (vii) social security systems created by the Federal Government, States, and Municipalities. **Hong Kong:** Distributed by UBS Securities Asia Limited and/or UBS AG, Hong Kong Branch. **Singapore:** Distributed by UBS Securities Pte. Ltd. [MCI (P) 018/09/2015 and Co. Reg. No.: 198500648C] or UBS AG, Singapore Branch. Please contact UBS Securities Pte. Ltd., an exempt financial adviser under the Singapore Financial Advisers Act (Cap. 110); or UBS AG, Singapore Branch, an exempt financial adviser under the Singapore Financial Advisers Act (Cap. 110) and a wholesale bank licensed under the Singapore Banking Act (Cap. 19) regulated by the Monetary Authority of Singapore, in respect of any matters arising from, or in connection with, the analysis or document. The recipients of this document represent and warrant that they are accredited and institutional investors as defined in the Securities and Futures Act (Cap. 289). **Japan:** Distributed by UBS Securities Japan Co., Ltd. to professional investors (except as otherwise permitted). Where this document has been prepared by UBS Securities Japan Co., Ltd., UBS Securities Japan Co., Ltd. is the author, publisher and distributor of the document. Distributed by UBS AG, Tokyo Branch to Professional Investors (except as otherwise permitted) in relation to foreign exchange and other banking businesses when relevant. **Australia:** Clients of UBS AG: Distributed by UBS AG (Holder of Australian Financial Services License No. 231087). Clients of UBS Securities Australia Ltd: Distributed by UBS Securities Australia Ltd (Holder of Australian Financial Services License No. 231098). Clients of UBS Wealth Management Australia Ltd: Distributed by UBS Wealth Management Australia Ltd (Holder of Australian Financial Services Licence No. 231127). This Document contains general information and/or general advice only and does not constitute personal financial product advice. As such, the Information in this document has been prepared without taking into account any investor's objectives, financial situation or needs, and investors should, before acting on the Information, consider the appropriateness of the Information, having regard to their objectives, financial situation and needs. If the Information contained in this document relates to the acquisition, or potential acquisition of a particular financial product by a 'Retail' client as defined by section 761G of the Corporations Act 2001 where a Product Disclosure Statement would be required, the retail client should obtain and consider the Product Disclosure Statement relating to the product before making any decision about whether to acquire the product. The UBS Securities Australia Limited Financial Services Guide is available at: www.ubs.com/ecs-research-fsg. **New Zealand:** Distributed by UBS New Zealand Ltd. The information and recommendations in this publication are provided for general information purposes only. To the extent that any such information or recommendations constitute financial advice, they do not take into account any person's particular financial situation or goals. We recommend that recipients seek advice specific to their circumstances from their financial advisor. **Korea:** Distributed in Korea by UBS Securities Pte. Ltd., Seoul Branch. This document may have been edited or contributed to from time to time by affiliates of UBS Securities Pte. Ltd., Seoul Branch. **Malaysia:** This material is authorized to be distributed in Malaysia by UBS Securities Malaysia Sdn. Bhd (Capital Markets Services License No.: CMSL/A0063/2007). This material is intended for professional/institutional clients only and not for distribution to any retail clients. **India:** Prepared by UBS Securities India Private Ltd. (Corporate Identity Number U67120MH1996PTC097299) 2/F, 2 North Avenue, Maker Maxity, Bandra Kurla Complex, Bandra (East), Mumbai (India) 400051. Phone: +912261556000. It provides brokerage services bearing SEBI Registration Numbers: NSE (Capital Market Segment): INB230951431, NSE (F&O Segment) INF230951431, NSE (Currency Derivatives Segment) INE230951431, BSE (Capital Market Segment) INB010951437; merchant banking services bearing SEBI Registration Number: INM000010809 and Research Analyst services bearing SEBI Registration Number: INH000001204. UBS AG, its affiliates or subsidiaries may have debt holdings or positions in the subject Indian company/companies. Within the past 12 months, UBS AG, its affiliates or subsidiaries may have received compensation for non-investment banking securities-related services and/or non-securities services from the subject Indian company/companies. The subject company/companies may have been a client/clients of UBS AG, its affiliates or subsidiaries during the 12 months preceding the date of distribution of the research report with respect to investment banking and/or non-investment banking securities-related services and/or non-securities services. With regard to information on associates, please refer to the Annual Report at: http://www.ubs.com/global/en/about_ubs/investor_relations/annualreporting.html

The disclosures contained in research documents produced by UBS Limited shall be governed by and construed in accordance with English law.

# Contact information

**UBS Limited**
1 Finsbury Avenue
London

Tel: +44-207-567 8000

www.ubs.com