# DatA414 Tutorial 3: Decision Trees

## 1. Decision tree absolute basics

Given the tree below, what would the input $\mathbf{x} = [3, 2, 4, 1]^\top$ be classified as given that $x_1$ is `sepal length`, $x_2$ is `sepal width`, $x_3$ is `petal length` and $x_4$ is `petal width`?
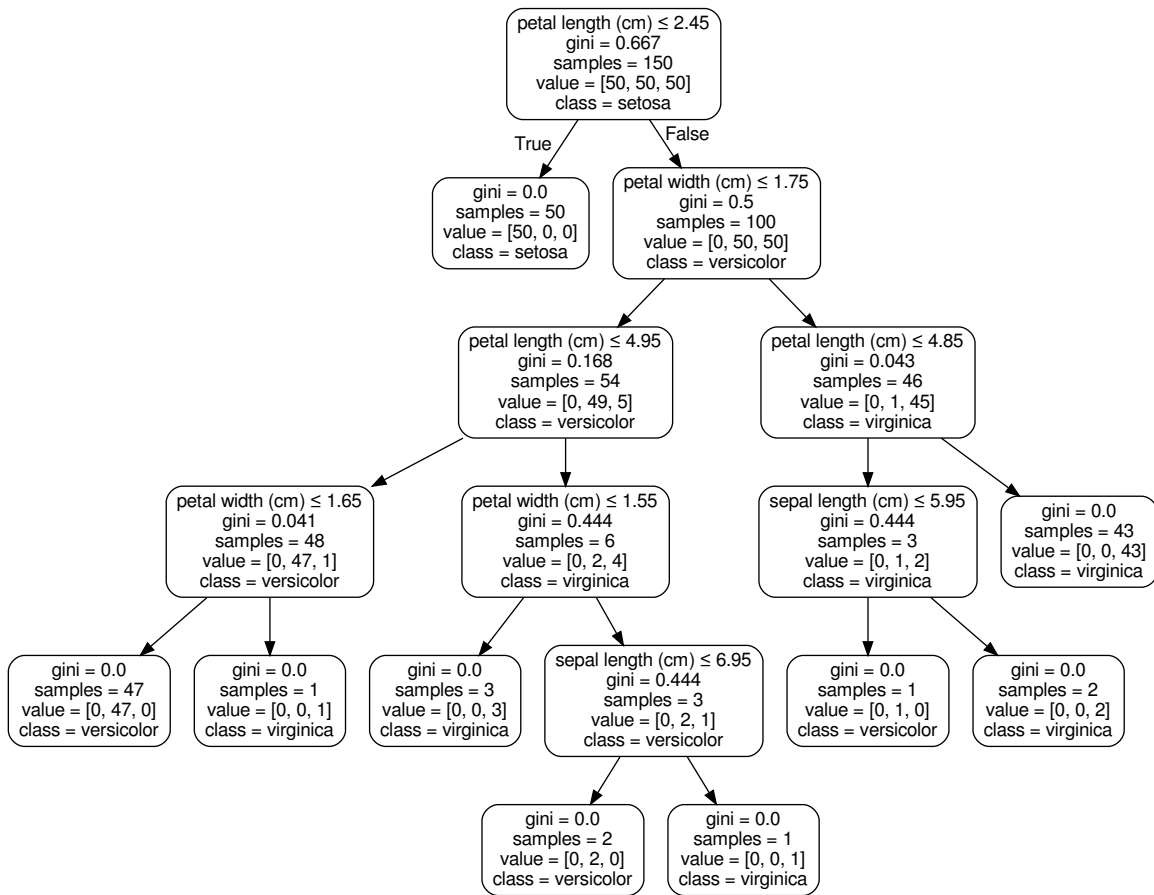


Figure 1: A decision tree trained on the Iris dataset.

## 2. Regression trees absolute basics

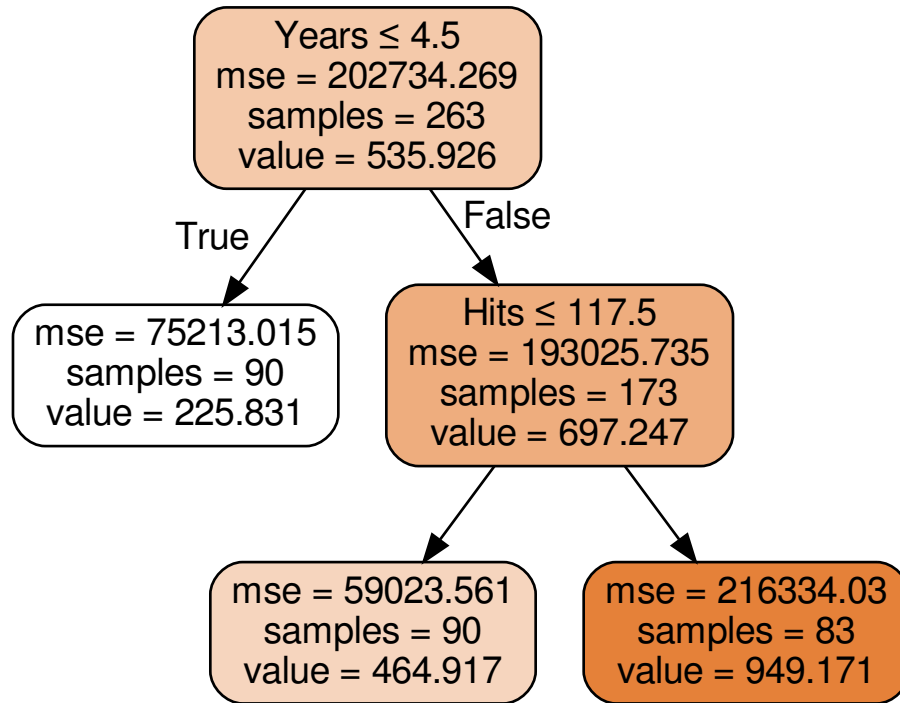Below is a regression tree trained on the hitters dataset.
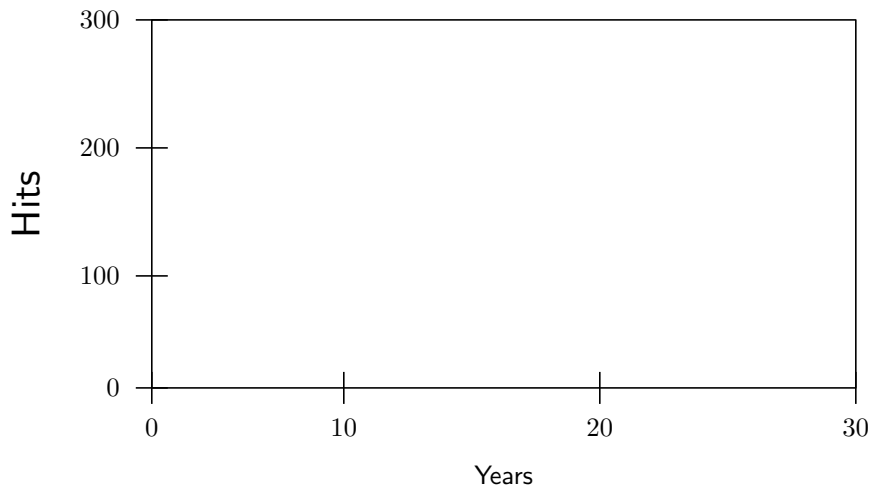


Figure 2: A regression tree trained on the hitters data.

Draw the decision boundary and prediction in each region for this tree on the feature space below:

## 3. Categorical features in regression trees

We want to train a regression tree that takes a person's occupation and predicts their salary. We consider three possible occupations: `artist`, `engineer` and `scientist`. We encode the occupation as a single feature $x_1$ as follows:

$$x_1 = \begin{cases} 0 & \text{if } \texttt{artist} \\ 1 & \text{if } \texttt{engineer} \\ 2 & \text{if } \texttt{scientist} \end{cases}$$

We have seen before that such an encoding is not a good idea when using linear regression. But what about regression trees?

Let's say we know that artists and scientists both earn the same average salary of R40 000 per month, while engineers on average earn R60 000. By hand, draw a regression tree that takes occupation as input, encoded as $x_1$, and predict these average salaries.

Is this $x_1$-encoding problematic or not, and how is this different or the same as the problem faced with linear regression?

## 4. Decision tree misclassification rate and Gini index

Consider a dataset with 400 examples of class $C_1$ and 400 of class $C_2$. Let tree $A$ have 2 leaves with class distributions:

| Tree $A$ | $C_1$ | $C_2$ |
|---|---|---|
| Leaf 1 | 100 | 300 |
| Leaf 2 | 300 | 100 |

and let tree $B$ have 2 leaves with class distribution:

| Tree $B$ | $C_1$ | $C_2$ |
|---|---|---|
| Leaf 1 | 200 | 400 |
| Leaf 2 | 200 | 0 |

What is the misclassification rate for both trees? Which tree is more pure according to the Gini index?

## 5. Decision tree construction

| $x_1$ | $x_2$ | $x_3$ | $y$ |
|-------|-------|-------|-----|
| 0     | 0     | 0     | 0   |
| 1     | 0     | 1     | 1   |
| 1     | 1     | 0     | 1   |
| 1     | 1     | 1     | 0   |

The dataset above contains three binary features with a binary target variable $y$. We want to build a decision tree which classifies $y$ as 0 or 1. Draw the tree that would be learned by the greedy tree growing algorithm giving a zero training loss.