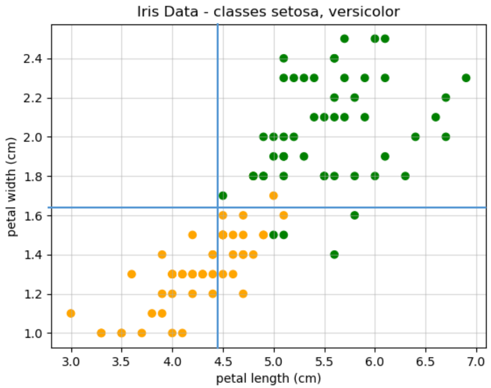
Entropy

- Since DTs split the data space s.t. classes are well separated, an impurity measure is needed
- **Entropy** and **Gini** Impurity are two usual choices, but much more measures were developed
- Direct reduction of the **classification error is not used**



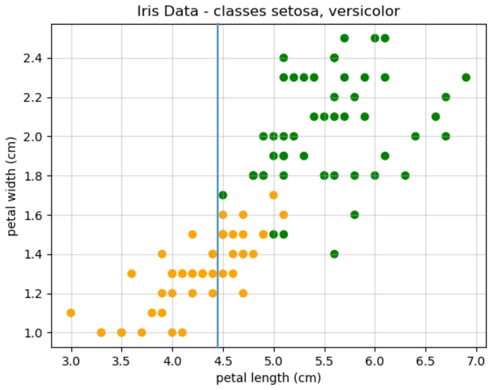
Use this split?

Use this split?

[Unifying attribute splitting criteria of decision trees by Tsallis entropy | IEEE Conference Publication | IEEE Xplore]

Entropy

- After splitting the data space, the **impurity measure is caluclated** for all **resulting partitions**



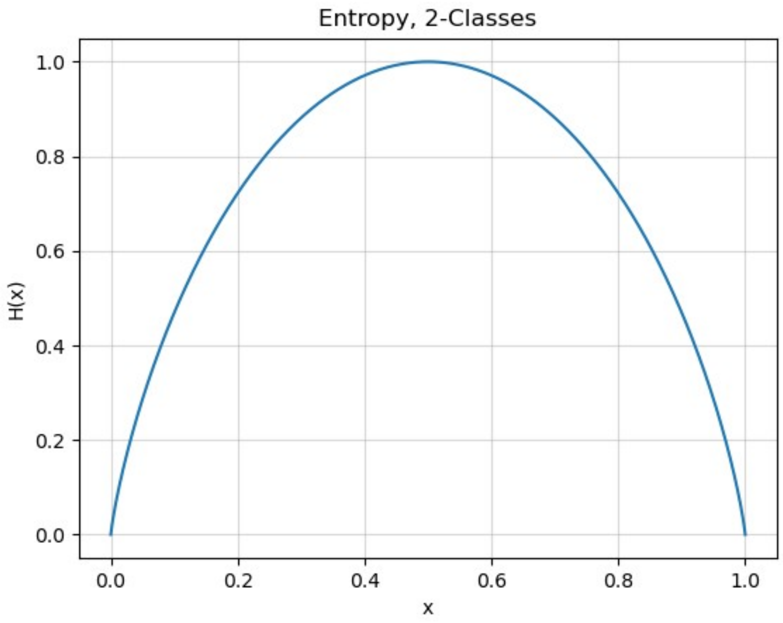
$$H = -\sum_i p(c_i) \log_2 p(c_i)$$

$p(c_i)$: relative frequency of class c_i inside the partition

Using split X1 > 4.4, we have 21 class1 and 50 class2 data points
p(c1) = 0.296, p(c2) = 0.704
H = -1*((0.296)*log2(0.296) + (0.704)*log2(0.704)) = 0.876

Entropy

- Entropy is **maximal** when **uncertainty is maximal**
- For a 2-class problem max(H) = 1 (in general H **not bounded**!)
- Maximum uncertainty is given, when classes are **evenly distributed**



Entropy

- Entropy gives us the **impurity of each data space partition**
- When comparing two or more possible splits, it **does not tell us which one is better**
- For split comparison **information gain is used**

$$\Delta H = H_p - \left[\frac{n_1}{n} H_1 + \frac{n_2}{n} H_2\right]$$

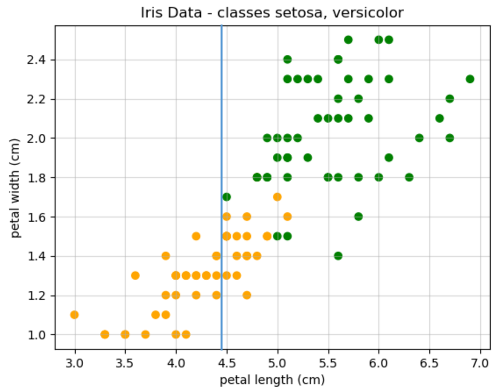
H_p: Entropy of parent node, H_1/H_2 : Entropy of partition 1/2,*

n_1/n_2: number of samples in partition 1/2, n: number of samples in the parent partition

FH University of Applied Sciences TECHNIKUM WIEN

*You'll often read partition left/right instead of 1/2

Entropy



Decision/Split: X1 <= 4.4

Parent partition
n = 101
#c1 = 50, #c2 = 50
P(c1) = 0.5, p(c2) = 0.5
H = 1

Decision = True (partition1):
n1 = 30
#c1 = 0, #c2 = 29
p(c1) = 0, p(c2) = 1
H1 = 0

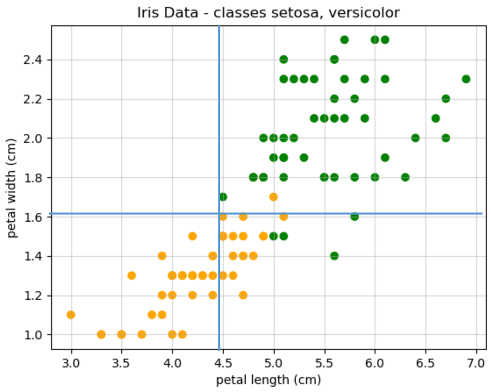
Decision = False (partition2)
n2 = 71,
#c1 = 21, #c2 = 50,
p(c1) = 0.296, p(c2) = 0.704
H2 = 0.876

Information Gain
n1/n = 0.297
n2/n = 0.703
$\Delta H = 1 – (0.297*0 + 0.703*0.876)$
$\Delta H = 0.384$

FH University of Applied Sciences TECHNIKUM WIEN

Entropy

- **Balancing by relative partition size** prevents extreme splits at the end of the value spectrum
- Information gain calculations continue for each path down the tree
- For every split, the node that is currently worked on is used as a reference node (= the parent node)



Decision/Split: X1 <= 4.4
$\Delta H = 0.384$

Decision/Split: X2 <= 1.6
$\Delta H = 0.7$

Parent Node = whole data space
H = 1

Gini Impurity/Index

- **Gini Impurity/Index** is an alternative to entropy
- Slightly **faster** to compute, isolates most frequent class in ist own branch, entropy produces more balanced trees buit **there‘s no big difference**
- Tree building/comparing splits with gini works similar to entropy

$$G = 1 - \sum_i p(c_i)^2$$

$p(c_i)$: relative frequency of class i inside a partition