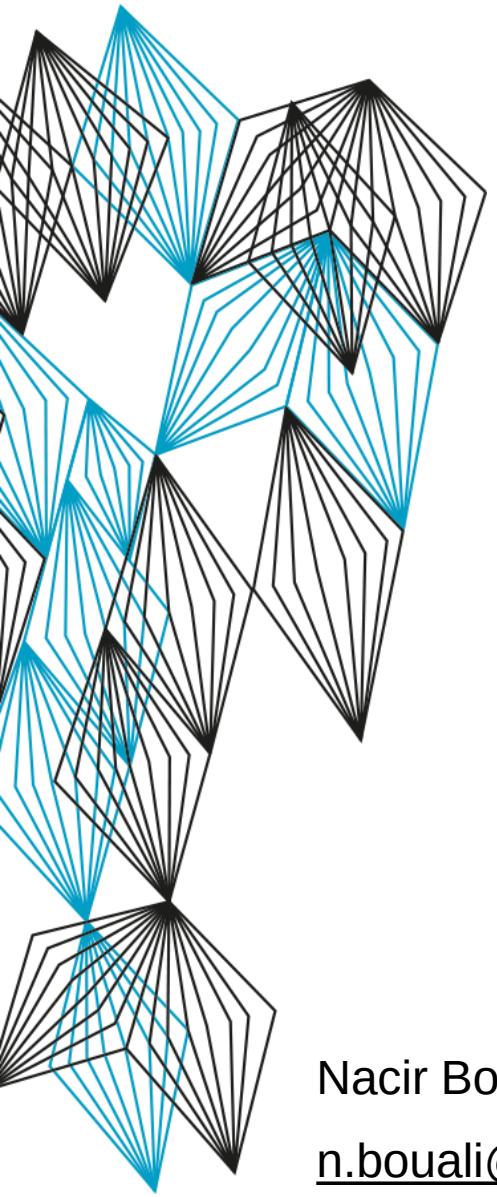


ARTIFICIAL INTELLIGENCE & CYBER SECURITY

MACHINE LEARNING
PRUNING DECISION TREES

Nacir Bouali

n.bouali@utwente.nl

UNIVERSITY OF TWENTE.

PRUNING DECISION TREES

- Decision trees have a major drawback
  - They can grow too deep
  - They overfit
- How do we deal with a model that overfits?
  - Remove splits that are not « statistically significant »

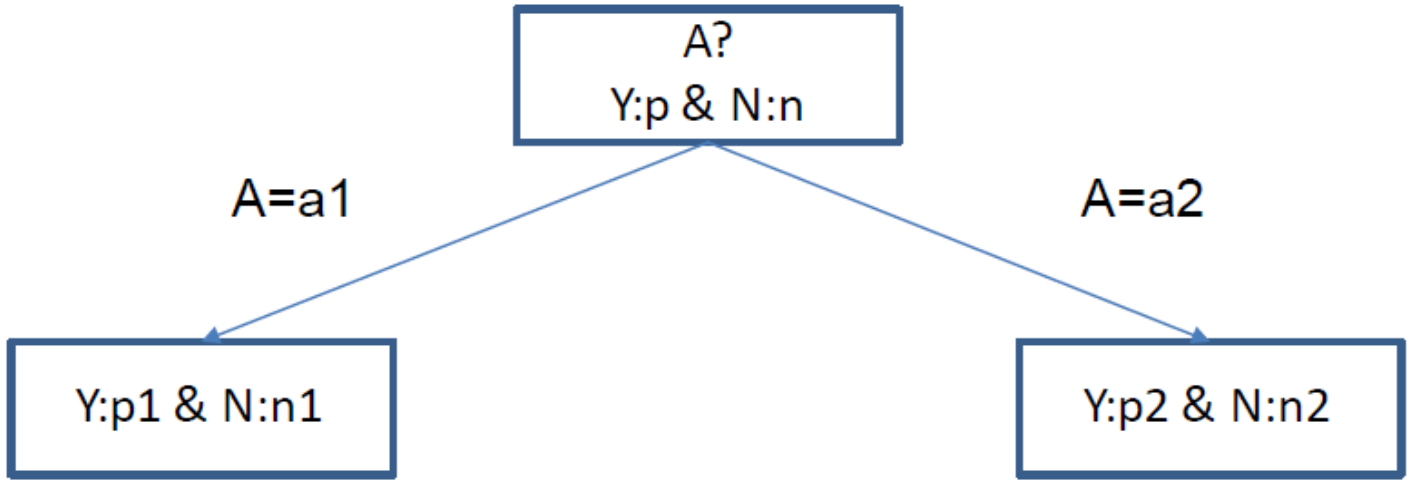
PRUNING DECISION TREES

- Pruning a decision tree while building it won't work.
- Famous XOR example.

A	B	Class	
a1	b1	Y:50	N:0
a1	b2	Y:0	N:50
a2	b1	Y:0	N:50
a2	b2	Y:50	N:0

- A and B have no information gain, but in combination they can separate the classes.

BOTTOM-UP PRUNING



No improvement then:

$p/(p+n)=p1/(p1+n1)=p2/(p2+n2)$

$n/(p+n)=n1/(p1+n1)=n2/(p2+n2)$

BOTTOM-UP PRUNING

- $\frac{p}{p+n} = \frac{p1}{p1+n1} \rightarrow p1 = p * \frac{p1+n1}{p+n}$
- Define $\widehat{p1} = p * \frac{p1+n1}{p+n}$, similar for n1, p2 and n2.
- Calculate:

$$\Delta = \frac{(p1-\widehat{p1})^2}{\widehat{p1}} + \frac{(n1-\widehat{n1})^2}{\widehat{n1}} + \frac{(p2-\widehat{p2})^2}{\widehat{p2}} + \frac{(n2-\widehat{n2})^2}{\widehat{n2}}$$

BOTTOM-UP PRUNING

- Δ is χ^2distributed with 1 degree of freedom. In general *d-1* degrees of freedom, with *d* number of splits.
- Small values for Δ imply no rejection of null hypothesis (H_0: no improvement) and hence pruning of tree.

CONCLUSION

- How decision trees learn from the data = fit the model.
- Problem of overfitting in decision trees.
- Evaluating decision trees on imbalanced datasets.
- Pruning statistically insignificant splits.