Quiz: Decision Trees

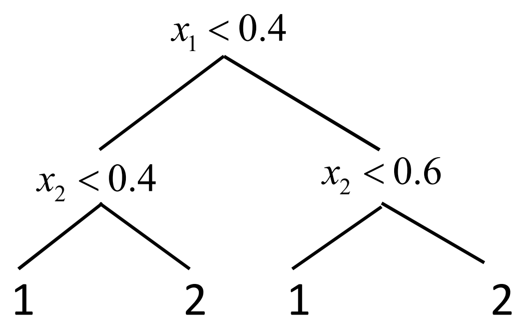
Lecture series „Machine Learning“

Niels Landwehr

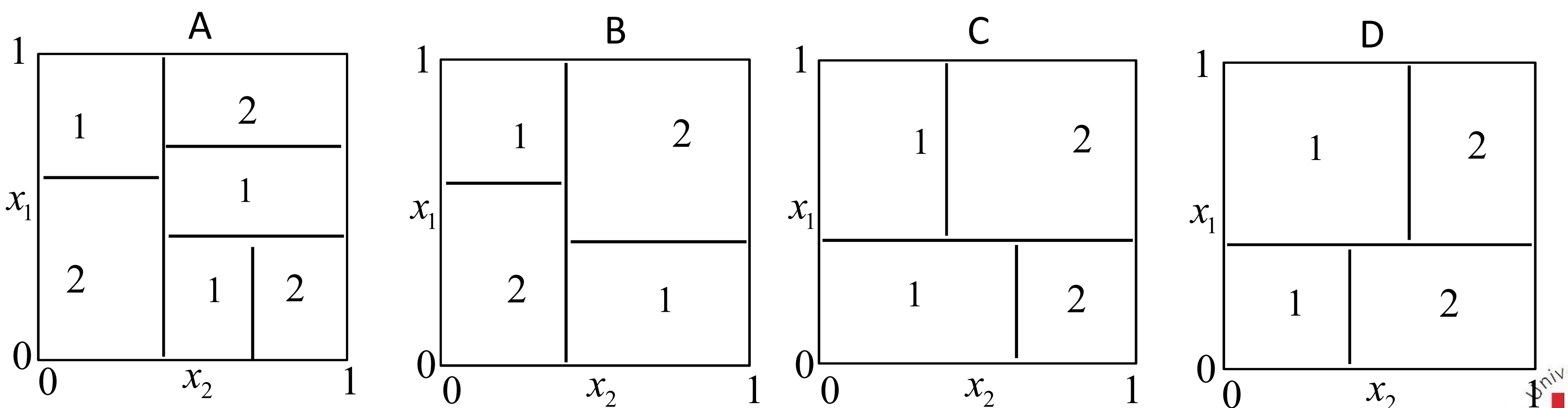
Research Group „Data Science“
Institute of Computer Science
University of Hildesheim

Quiz: Tree Partitions

- Assume a classification problem with two input attributes $x_1, x_2 \in \mathbb{R}$ and classes $\{1,2\}$. The following decision tree implements a function $f : \mathbb{R}^2 \to \{1,2\}$:

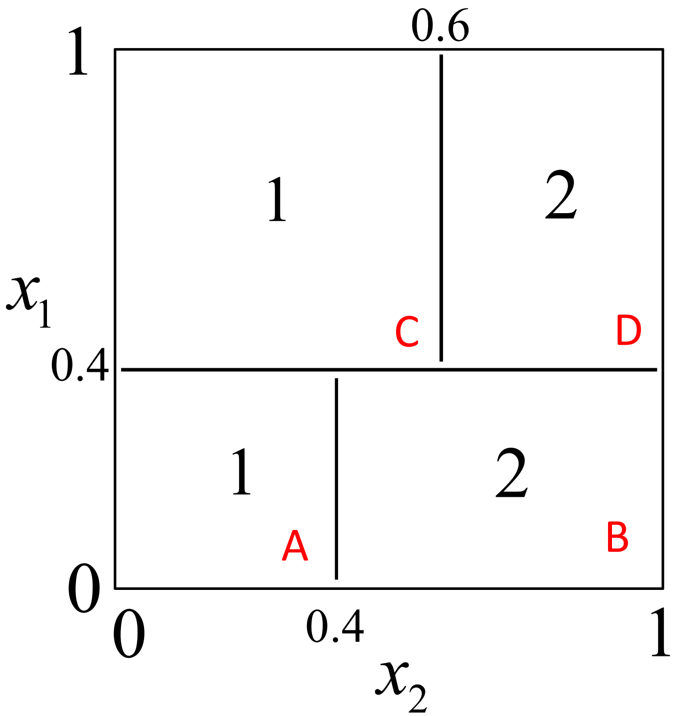


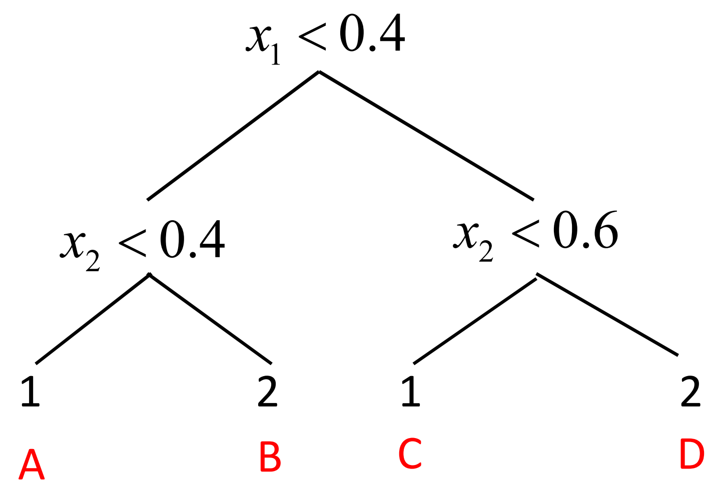
- Question: which of the following partitions is the one represented by the tree?



Solution: Tree Partitions

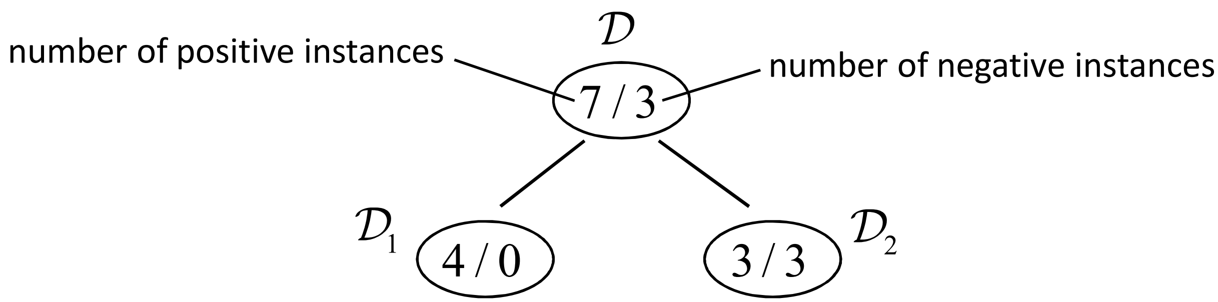
- **Solution**: Partition D





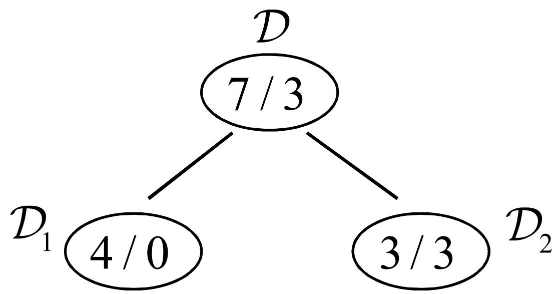
Quiz: Gini Index

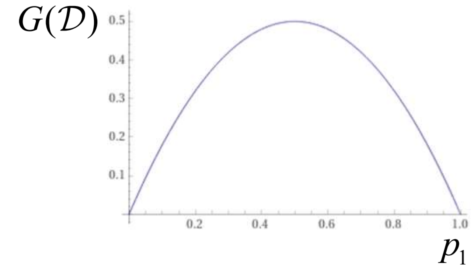
- Consider the following split for a decision tree for a binary classification problem:



- What is the Gini index for this split?
  - It is 0
  - It is 0.05
  - It is 0.12
  - It is 0.43
  - It is 0.67

Solution: Gini Index





- **Solution**: The Gini index is 0.12
  - We first compute the Gini impurity of the root data set $\mathcal{D}$:

$$G(\mathcal{D}) = 1 - \sum_{t=1}^{T} p_t^2 = 1 - 0.7^2 - 0.3^2 = 0.42$$

  - Then we need the Gini impurities of $\mathcal{D}_1$ and $\mathcal{D}_2$: these are 0 and 0.5
  - The Gini index is then

$$GI(\mathcal{D}, \mathcal{D}_1, \mathcal{D}_2) = G(\mathcal{D}) - \sum_{r=1}^{2} \frac{|\mathcal{D}_r|}{|\mathcal{D}|} G(\mathcal{D}_r) = 0.42 - \frac{4}{10} 0 - \frac{6}{10} 0.5 = 0.12$$

Quiz: Decision Tree and Linear Regression

- Assume a regression problem with two input attributes $x_1, x_2 \in \mathbb{R}$
- Assume a linear regression model $f_{\boldsymbol{\theta}} : \mathbb{R}^2 \to \mathbb{R}$ given by $f_{\boldsymbol{\theta}}(\mathbf{x}) = \theta_1 x_1 + \theta_2 x_2$

- Question: Is there a regression tree $f : \mathbb{R}^2 \to \mathbb{R}$ with univariate splits of the form $x_i < C$ that implements the same function?
  – Yes, always
  – Yes, but only if we limit the input to a finite (multi-dimensional) interval
  – That depends on the $\boldsymbol{\theta}$: for some there is an equivalent tree, for some not
  – An equivalent tree only exists for the trivial case $\boldsymbol{\theta} = \mathbf{0}$
  – No, never

Solution: Decision Tree and Linear Regression

- **Solution**: only for the trivial case $\boldsymbol{\theta} = \mathbf{0}$
  – If $\boldsymbol{\theta} \neq \mathbf{0}$, the output domain of the linear model $f_{\boldsymbol{\theta}}(\mathbf{x})$ is continuous: the set $\{f_{\boldsymbol{\theta}}(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^2\}$ is not finite
  – For a (finite) regression tree $f(\mathbf{x})$, the number of possible outputs $|\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^2\}|$ is bounded by the number of leaves
  – This holds even when we restrict the input to a finite interval
  – However, on a finite interval the tree can approximate the linear model arbitrarily closely, if the tree is large enough

Quiz: Decision Tree and Logistic Regression

- Assume a classification problem with two inputs $x_1, x_2 \in \mathbb{R}$ and two classes $\{0,1\}$

- Assume a logistic regression model $f_{\boldsymbol{\theta}} : \mathbb{R}^2 \rightarrow \{0,1\}$ given by

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \begin{cases} 1: & \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2) \geq 0.5 \\ 0: & \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2) < 0.5 \end{cases}$$

- Question: Is there a decision tree $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ with univariate splits of the form $x_i < C$ that implements the same classifier?
  - Yes, always
  - Yes, but only if we limit the input to a finite (multi-dimensional) interval
  - That depends on the $\boldsymbol{\theta}$: for some there is an equivalent tree, for some not
  - An equivalent tree only exists for the trivial case $\boldsymbol{\theta} = \mathbf{0}$
  - No, never

Solution: Decision Tree and Linear Regression

- **Solution**: that depends on the $\boldsymbol{\theta}$
  - A decision tree with splits $x_i < C$ can only represents axis-parallel decision boundaries
  - The logistic regression model has an axis-parallel decision boundary if and only if either $\theta_1 = 0$ or $\theta_2 = 0$
  - So for most logistic regression models it does not work, but for example it works for $\boldsymbol{\theta} = (2,0,1)$: class one is predicted if

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 > 0$$

$$2 + x_2 > 0$$

$$x_2 > -2$$

  - Equivalent tree would be