ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

HS 2023

Prof. R. Wattenhofer
Benjamin Estermann

Computational Thinking
Exercise 10

1 Global Minimum

You want to find the global minimum of f using gradient descent, where $f = 3x^4 - 4x^3 - 12x^2 + 4$

a) Given a small enough learning rate, for what range of initial values x_0 will gradient descent converge to the global minimum?

b) Let $x_0 = 3$. What is the optimum learning rate to reach the global minimum with the least number of steps?

c) Does Newton's method use the optimum learning rate? Why (not)?

d)* What about if $f = ax^2 + bx + c$ with $a > 0$ and we have an arbitrary starting point x_0?

2 Logistic Regression & XOR

We want to learn the "XOR" function with logistic regression. Our input space is $\mathcal{X} = \{0,1\}^2$ and our output space is $\mathcal{Y} = \{0,1\}$ and we want to learn the mapping

$$(x_1, x_2) \mapsto x_1 \oplus x_2$$

a) Why can logistic regression not learn "XOR"?

b) Show that logistic regression can learn "XOR" by manually adding features.

c) How about "AND", "OR", "NOT AND"? Can logistic regression learn these?

d) Show that "hierarchical" logistic regression with 2 layers can learn "XOR". What does this remind you of?

e) How about a decision tree, can it learn "XOR"?

3 Gini Impurity

Definition 10.1 (Classification splitting criterion: Gini Impurity)**.** *For node v containing samples D_v from k classes, the gini measure of impurity is defined as:*

$$G = 1 - \sum_{i=1}^{k} p_i^2$$

where

$$p_i = \frac{\{\boldsymbol{x} \in D_v \mid f(\boldsymbol{x}) = i\}}{|D_v|}$$

is the fraction of samples within D_v that belongs to class i.

Take a look at this data!

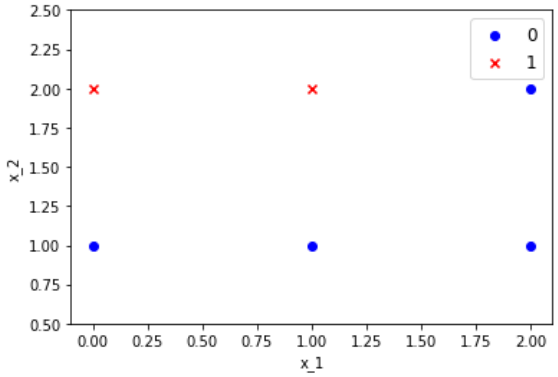

Figure 1: Some binary data

a) Construct an optimal decision tree (requiring the minimum number of splits).

b) Show that we find an optimal decision tree by using the CART loss function with Gini impurity.

c) Give an example dataset, where CART with Gini does not find an optimal decision tree.