Astronomy & Astrophysics manuscript no. project2
October 12, 2022

Classification and regression, from linear and logistic regression to neural network

Johan Mylius Kroken[1,2], and Nanna Bryne[1,2]

[1] Institute of Theoretical Astrophysics (ITA), University of Oslo, Norway
[2] Center for Computing in Science Education (CCSE), University of Oslo, Norway

October 12, 2022 GitHub repo link: https://github.com/Johanmkr/FYS-STK4155colab/tree/main/project2

ABSTRACT

ABSTRACT

rephrase this (\rephrase{...})
check if this is correct (\checkthis{...})
* *comment (\comment{...})*
add filler text (\fillertext)
for when you are lost (\wtf)

Nomenclature

1. Introduction

2. Theory

In linear regression we have the famous assumption that

$$y_i = f(\mathbf{x}_i) + \varepsilon_i \simeq \mathbf{x}_i^\intercal \boldsymbol{\beta} + \varepsilon_i, \tag{1}$$

where f is a continous funtion of $\mathbf{x}$. If we now were to allow said function to represent discrete outputs, we would benefit from moving on to *logsitic* regression.

* *Maybe move to intro?*

add filler text smooth transition to steepest descent

2.1. Gradient descent

Gradient descent (Hjorth-Jensen 2021)

add filler text

The most basic concept is that of steepest descent. In order to find a minimum of a function $f(\mathbf{x})$ (allow multivariability of $\mathbf{x}$), we follow the steepest descent of that function, i.e. the direction of the negative gradient $-\nabla f(\mathbf{x})$. We thus have an iterative scheme to find minima:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \nabla f(\mathbf{x}_k), \tag{2}$$

where γ_k may be referred to as the step length or learning rate, which

2.1.1. Plain gradient descent (GD)

2.1.2. Stochastic gradient deprojescent (SGD)

2.2. Tuning

* *Something about hyper parameters λ and γ*

3. Conclusion

Code availability

The code is available on GitHub at https://github.com/Johanmkr/FYS-STK4155colab/tree/main/project2.

References

Hjorth-Jensen, M. 2021, Applied Data Analysis and Machine Learning (Jupyter Book)