

Data Analysis and Machine Learning: Linear Regression and more Advanced Regression Analysis

Morten Hjorth-Jensen^{1,2}

Department of Physics, University of Oslo¹

Department of Physics and Astronomy and National Superconducting Cyclotron Laboratory, Michigan State University²

Oct 11, 2017

© 1999-2017, Morten Hjorth-Jensen. Released under CC Attribution-NonCommercial 4.0 license

Regression analysis, overarching aims

Regression modeling deals with the description of the sampling distribution of a given random variable y varies as function of another variable or a set of such variables $\hat{x} = [x_0, x_1, \dots, x_p]$. The first variable is called the **dependent**, the **outcome** or the **response** variable while the set of variables \hat{x} is called the independent variable, or the predictor variable or the explanatory variable.

A regression model aims at finding a likelihood function $p(y|\hat{x})$, that is the conditional distribution for y with a given \hat{x} . The estimation of $p(y|\hat{x})$ is made using a data set with

- ▶ n cases $i = 0, 1, 2, \dots, n - 1$
- ▶ Response (dependent or outcome) variable y_i with $i = 0, 1, 2, \dots, n - 1$
- ▶ p Explanatory (independent or predictor) variables $\hat{x}_i = [x_{i0}, x_{i1}, \dots, x_{ip}]$ with $i = 0, 1, 2, \dots, n - 1$

The goal of the regression analysis is to extract/exploit relationship between y_i and \hat{x}_i in or to infer causal dependencies,

General linear models

more text to come