Modern Regression

Jean-Luc Bouchot

School of Mathematics and Statistics
Beijing Institute of Technology
jlbouchot@bit.edu.cn

Spring 2021

Outline

Remark 2.1

So far, we have introduced the basics of regression from a statistical point of view. Let us, until the end of the semester, extend these results to more general theoretical results and practical algorithms, including:

- Multiple linear regression: linear model in d dimensions
- Multiple linear regression: polynomial models in 1D and curve fitting
- Multiple linear regression: general models
- Model selection and regularization by cross validation
- Logistic regression and classification
- Neural networks: perceptron, backpropagation
- Deep models

Remark 2.2

Let us re-introduce the linear model studied so far:

$$f_\theta(x_i) = \alpha + \beta x_i = \mathbf{x}_i \theta,$$

where $\theta = [\alpha, \beta]^T$ is the (column) vector for parameter of the model and $\mathbf{x}_i = [1, x_i]$ corresponds to the samples.

Definition 2.1

Multiple linear regression *corresponds to a linear regression model in which we have more than one predictors or independent variables. Hence, we have*

$$f_\theta(\mathbf{x}) = \mathbf{x}\theta,$$

where $\theta = [\theta_0, \theta_1, \cdots, \theta_d]^T$ *is the* $d+1$ *dimensional vector of parameters and* $\mathbf{x} \in \mathbb{R}^{1\times d}$ *is the* d *dimensional row vector of the sample.*
A sample may also be called an **individual**. *A dimension of the individual is sometimes called a* **feature**.

Example 2.1

It is hypothesized that lobsters continue to grow for as long as they live. One could, for instance, hypothesized that the age of a lobster is related (linearly) to its size (length) and weight. In this case, we have

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2,$$

with x_1 its weight (say in hundreds of grams) and x_2 its size (say in decimeters). We have

$$y = \mathbf{x}\theta.$$

Definition 2.2

We may have n samples (or individuals) which we can use to estimate our parameters and / or evaluate our model. In this case, we might create a **data matrix***:*

$$X = [\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_n^T]^T \in \mathbb{R}^{n\times(d+1)}.$$

Remark 2.3

It is important to keep some clear ideas on the data matrix:

- One row of the data matrix corresponds to one individual.
- One column of the data matrix corresponds to one feature.
- Some people prefer to deal with the data matrix as the transposed version of the one introduced here. Make sure you know what you are doing when using data matrices.
- $\mathbf{x}_1$ corresponds to the first individual in our data matrix / dataset. $\mathbf{x}_{i,j}$ corresponds to the j^{th} feature of the i^{th} sample. x_i corresponds to the i^{th} feature.
- Be careful! Some people count the bias (feature associated to θ_0) as a feature. In this class, we will try to call features when we are actually measuring / computing something, i.e. they are d features numbered 1 to d and one bias (what would be the feature number 0).

Exercise 2.1

The height of a child might depend on the height of the parents, on some other factors. Write a multiple linear regression model and how the samples would be measured if the height depended on

1. the height of the parents only,
2. the height of the parents and the amount of fibers eaten by the mother during the pregnancy
3. the height of the parents, the amount of fibers eaten by the mother, and the number of cigarettes per day smoked by the mother.

Example 2.2

Polynomial features are in fact a linear model (with respect to the parameters):

$$f_\theta(x) = \mathbf{x}\theta$$

where $\theta \in \mathbb{R}^{d+1}$ and we artificially *lift* the univariate predictor variable x to a d dimensional vector $\mathbf{x} = [x_{j_i}]_{1 \leq i \leq d}$ and $j_i \in \mathbb{N}$.

Example 2.3

The calibration parameter can be fitted by a linear regression with polynomial features.