# Foundations of data science

Steffen Mæland, PierGianLuca Porta Mana

2023-04-27

# Contents

# Preface

Under construction

# Chapter 1

# Introduction

# Chapter 2

# Truth inference and probability inference

## 2.1 Statements, well-posed and ill-posed

Facts, hypotheses, questions, decisions – and data are communicated through language and sentences. You may say "well, data can be just numbers, they don't need to be communicated through sentences". But is that true?

I give you this number: "5". OK it's a number, but what's it about? what should you do with it? is that "data"? Instead, if I tell you: "The number of lectures in this course is 5" then I have given you a piece of information, a datum (even if it is actually false). Underlying any piece of information, hypothesis, or datum, there is always a statement that gives you the meaning and context of that datum.

In fact we face problems when those statements aren't clear. Suppose that a car model consumes 150~Wh/km and has a trunk capacity of two medium-sized suitcases; a second car model consumes 250~Wh/km and has a trunk capacity of five medium-sized suitcases. Someone asks you: "which car is better?". Well, it isn't clear how you should answer; what does "better" mean? If it refers to consumption, then the first car is "better". If it refers to cargo space, then the second car is "better". If it refers to a combination of these two characteristics, or to something else, then we simply can't answer. Here we have a problem with querying and giving data, because the statement underlying such query is not clear. We say that statement is not **well-posed**, or that it is **ill-posed**.

This may seem an obvious discussion to you. Yet you'd be surprised by how often unclear statements appear in scientific papers about data engineering! Not seldom we find discussions and disagreements that actually come from unclear underlying statements, that two parties interpret in different ways.

As a data engineer, you'll often have the upper hand if you are on the lookout for ill-posed statements. Whenever you face an important question, or you're given an important piece of information, or you must provide an important piece of information, always take a little time to examine whether the question or information is actually well-posed.

# Chapter 3

# Literature

Here is a review of existing methods.