

# Sketching: A framework for compressive learning

Leo Davy, Martin Gjorgjevski, Aleksandr Pak

March 2022

## 1 Introduction

Dealing with large and high dimensional datasets can be problematic from computational perspective. Indeed, most of the classical optimization algorithms require all of the data to be loaded in memory and to be looped through multiple times which can be very prohibitive for practical applications. A framework proposed to deal with this issue is Sketching, where carefully designed transformations of the data lead to massive compression, i.e. the data is represented as a single vector, and optimization schemes are constructed using the compressed data.

Sketching is designed for unsupervised problems such as Principal Component Analysis (PCA), k-means clustering and Gaussian Mixture Models (GMMs) but it is possible to use it for supervised problems as well.

Many analogies with compressed sensing are available. The method can be understood as optimization over the space of probability distributions where certain parametric constraints are imposed on the model. There are theoretical guarantees for excess risk control under an assumption called Lower Restricted Isometry Property (LRIP).