

# Sequential Data Analysis with Hidden Markov Models

**Introduction.** We covered clustering on Day 2. Clustering assumes that the data are exchangeable, i.e., that the probability of a data set does not change if you change its order. Clearly data that have sequential structure, such as time series, are not exchangeable. The hidden Markov model (HMM) can be thought of as a clustering method for non-exchangeable data, and has similarities to K-means and Gaussian mixture models.

In this project, we extend the mixture model to the hidden Markov model and explore inference algorithms, applications, further extensions and limitations.

**Project description.** In this project, you extend the objective function from K-means to a simple HMM and learn how to optimize this new objective function with respect to the latent component assignments and the component means. After implementing the training procedure for the HMM, the rest of the project will be about analyzing stock data using the algorithm.

**Dataset.** The project will use real data from the stock market. The applied task will be to assign each day's trading to a latent component that will indicate what the direction of the market is expected to be the next day.