Lecture 3 Overview

Zi Jian Liu

Jan 29 2021

Regression and Classification

We started off covering Model based Interpretation of OLS Regression. We used the Gaussian Noise model and looked at the Joint log-likelihood, and found that the OLS estimator is the MLE and is optimal under the Gaussian Noise Model.

Why do we care about model based interpretations?

1. Optimal estimator

2. Confidence Intervals/p-values/statistical significance

3. Generative Story

4. Bayesian Inference

Feature Engineering

We first covered Linear regression with Basis Expansion and from linear to nonlinear:

1. Input X can be transformations of original features. (Handcrafted features)

2. Inputs can be interaction terms

3. Inputs can be basis expansions

4. Indicator functions of quantitative inputs, more generally, categorical data analysis

We then covered High Dimensional Data, where the dimensionality d is comparable to or greater than the sample size n. We covered two regularization methods, a two-stage procedure and a single-step method. We had a graphic illustration and solved for the Ridge estimator.

We covered a basic method of Model Selection, Data Splitting, which generalizes well and is theoretically and computationally simple, but wastes training data. The solution of which is to use J-fold cross validation for Model Selection.

Honor Code

I pledge my honor that this lecture summary is my own work and adheres to the guidelines in the instructions.

-Zi Jian Liu