

Seminar Writeup
Nan Tang

Anchor Regression: Heterogeneous Data Meets Causality

Dominik Rothenhäusler

2/1 3:30 - 4:30

Name: Qianqian Chen

Affiliation: First-year Master student in UW Biostatistics

Facts: Qianqian said her research interest is in survival analysis, such as the expected duration of time from the death of certain organism until failure of entire mechanical systems.

The seminar discussed a machine learning topic, the importance of anchor regression of heterogeneity on causality regression. Heterogeneity means the characteristics or features of population, sample or data set that are different. In the aspect of finding causality, heterogeneity is a good approach to seek for causal effect and causal parameter that yields reliable prediction. Anchor in training data set denotes exogeneous variable that captures changes in distribution in training and test data. Anchor regression optimizes the minimax risks under the set of perturbations, thereby causal predictions that are built under certain perturbations are more reliable. The application of anchor regression includes gene expression (where the speaker is working on). In the gene expression case, the dependent variable is expression of one selected gene, independent variable is expression of other genes, and the anchor is the genotyping principal components and the peer factors. Speaker also mentioned that there is always a tradeoff between low predictive accuracy caused by causal model, and unreliability caused by non-causal model.