W207– Applied Machine Learning

Cornelia Paulik, PhD

School of Information

UC Berkeley

Linear Regression – feature engineering

Announcements

- Next week: group, question, and dataset you plan to use for the final project
- AI/ML for Social Good: Real signs of a mass shooting (NYT article, data)

Last week

- General concepts of Linear Regression and Gradient Descent
- Predict a continuous outcome variable using the diabetes dataset.
- Didn’t have time to do the breakout room exercise ☹
- Introduction to TensorFlow2

Today's learning objectives

- Linear Regression but focus on Feature Engineering
  - Async + live session: different types of features (continuous, discrete, different distributions)
  - Async + live session: feature standardization
  - live session: closed form solution vs. gradient descent implementation
  - live session: what if panel data? (spatial and time series dimension)

- Application: Predict the growth rate of COVID-19 cases during the first wave of the pandemic (1. Linear_regression (features).ipynb)

- Breakout room exercise ☺

Paper review: Ilin et al. (2021)

- Define data used in the analysis
- How was the data split into training and test?
- Define modeling strategy (learning algorithms)
- Define evaluation metrics