Integrating Machine Learning into Undergraduate Statistics Education

Xuemao Zhang
East Stroudsburg University
The 2024 Electronic Conference on Teaching Statistics

June 10 - 13, 2024

Overview

- Background and Students Majors
- Statistics and Machine Learning Topics
- Course Evaluations
- Discussions
- Teaching Resources and References

Background

- ESU (East Stroudsburg University) is a small undergraduate university with about 5,000 students.
  - Statistics minor program (Mathematics) and data science certificate (Computer science).
- Statistics Courses:
  - Descriptive statistics can help us understand data better.
  - Statistical inferences focus on Confidence Intervals and Hypothesis Testing, often assuming normality.
- Machine learning focuses on predicting individual responses with models optimized for the lowest cross-validation (CV) errors.
- Current statistics knowledge is not sufficient for students' future careers.
- The Spring 2023 class (Math 402: Applied Statistical Methods) includes students from Mathematics, Computer Science, and Psychology majors.

Topics covered in the course

- Software: Python
- Statistics topics
  - Data Visualization
  - Analysis of Variance (ANOVA) for Design of Experiments
  - Analysis of Covariance (ANCOVA)
  - Analysis of Categorical Data
  - Regression Models
- Machine Learning topics
  - Machine Learning using Regression Models
  - Linear Model Regularization: Ridge Regression, LASSO, Principal Components Regression, Partial Least Squares
  - Classification: K-Nearest Neighbors, Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis

Course Evaluations

- Teaching focus on applications using Python
- Course Evaluations:
  - In-class quizzes
  - Homework assignments
  - Two major projects
  - Final Project

Discussions

- Integrating machine learning into statistics courses benefits students' career prospects.
- The focus on Supervised Machine Learning provides a balanced approach between traditional statistics and modern machine learning techniques.
- Future teaching: Determining the appropriate extent of AI integration in the curriculum.

Teaching Resources

- Chan, Stanley H. (2021). *Introduction to Probability for Data Science*. https://probability4datascience.com/
- Danielle Navarro and Ethan Weed (2021). *Learning Statistics with Python*. https://ethanweed.github.io/pythonbook/landingpage.html
- Wes McKinney (2016). *Python for Data Analysis* (3rd). https://wesmckinney.com/book/ https://wesmckinney.com/
- Stefanie Molin (2021). *Hands-On Data Analysis with Pandas: A Python data science handbook for data collection, wrangling, analysis, and visualization* (2nd). https://github.com/stefmolin/Hands-On-Data-Analysis-with-Pandas-2nd-edition
- James, G., Witten, D., Hastie, T., Tibshirani, R.and Taylor, J. (2023). *An Introduction to Statistical Learning: with Applications in Python*. Springer. https://www.statlearning.com/
- Python Cheat Sheets. https://www.utc.fr/~jlaforet/Suppl/python-cheatsheets.pdf
- Zhang, X (2024). Math 402 Lecture Notes https://github.com/esumath/Math402/.