Optimization, learning and high-dimensional macroscopic limits

Iris Chang
Previous works

- Logistic Regression: models binary outcomes
  - Ex: Patient diagnosis (Salahi et al., 2019)
  - Regularization: adding a penalty to prevent overfitting
- When $p$ fixed and $n \to \infty$, MLE has nice properties (e.g. unbiased)
- In high dimensions, properties break down for unregularized, $L_1$ and $L_2$ regularized logistic regression (Candès & Sur, 2018; Salahi et al., 2019)
Looking ahead + Goals

- Understand and replicate results from previous work
  - Generate data set and compute log reg estimates
  - Specifically relating to $L_2$ regularized log reg
- How can we mitigate these issues?
  - Bagging

Original Data

Bootstrapping

Aggregating

Bagging

(Lamyai, 2014)
Works Cited and Acknowledgements


Acknowledgements:

This work was carried out as a part of the 2023 DIMACS REU program at Rutgers University, supported by NSF grant CNS-2150186

Thank you to my mentor Pierre Bellec for his time and guidance throughout!