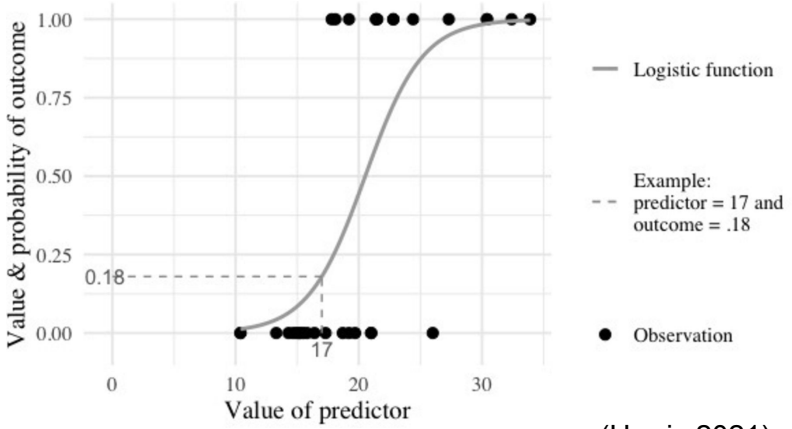
Optimization, learning and high-dimensional macroscopic limits

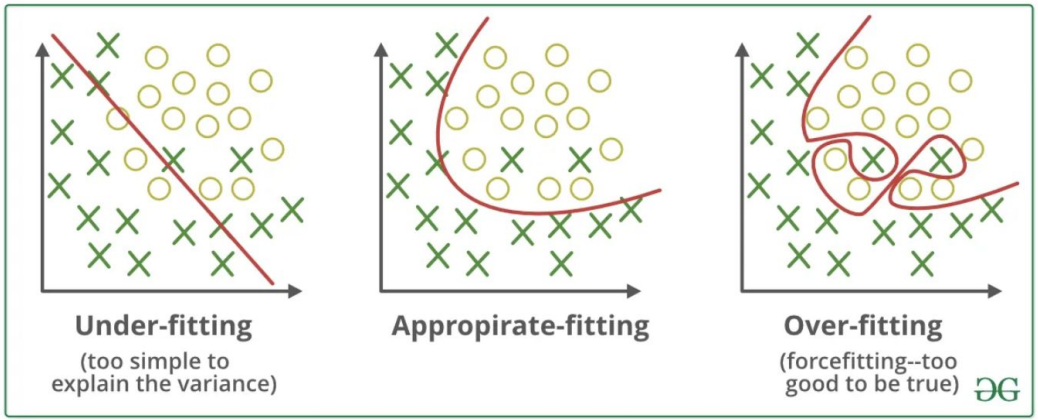
Iris Chang

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 → ∞, 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)



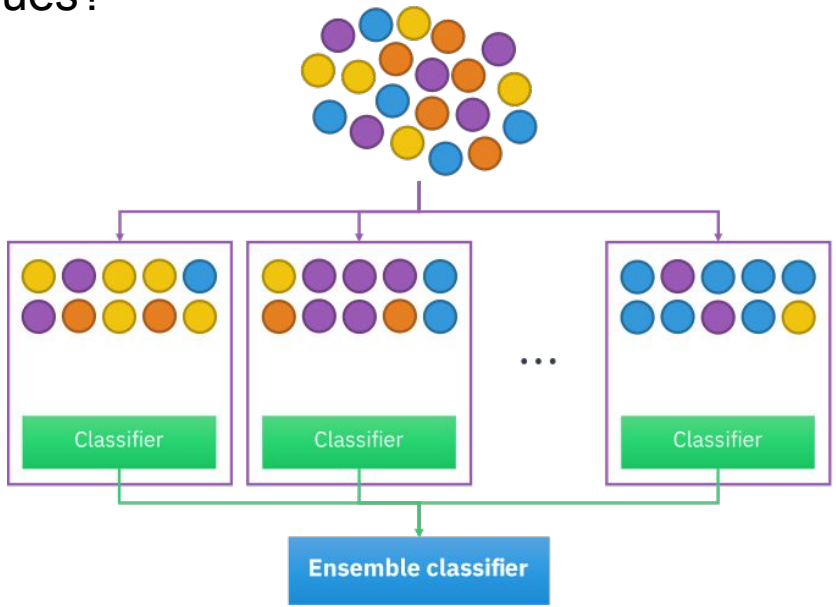
(Harris 2021)



(Karas 2023)

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



(Lamyai, 2014)

Works Cited and Acknowledgements

Emmanuel J Candès and Pragya Sur. (2018). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. arXiv preprint arXiv:1804.09753.

Harris JK. (2021). Primer on binary logistic regression. *Family Medicine and Community Health.* doi: 10.1136/fmch-2021-001290

Karas, P. (2023). *Under, over and appropriate fitting in logistic regression*. L1 (Lasso) and L2 (Ridge) regularizations in logistic regression. Medium . Retrieved from https://ai.plainenglish.io/l1-lasso-and-l2-ridge-regularizations-in-logistic-regression-53ab6c952f15 .

Lamyai, S. (2014). *An illustration for the concept of bootstrap aggregating*. Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Bootstrap_aggregating#/media/File:Ensemble_Bagging.svg.

Salehi, F., Abbasi, E., & Hassibi, B. (2019). *The Impact of Regularization on High-Dimensional Logistic Regression*. https://doi.org/ https://doi.org/10.48550/arXiv.1906.03761

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!