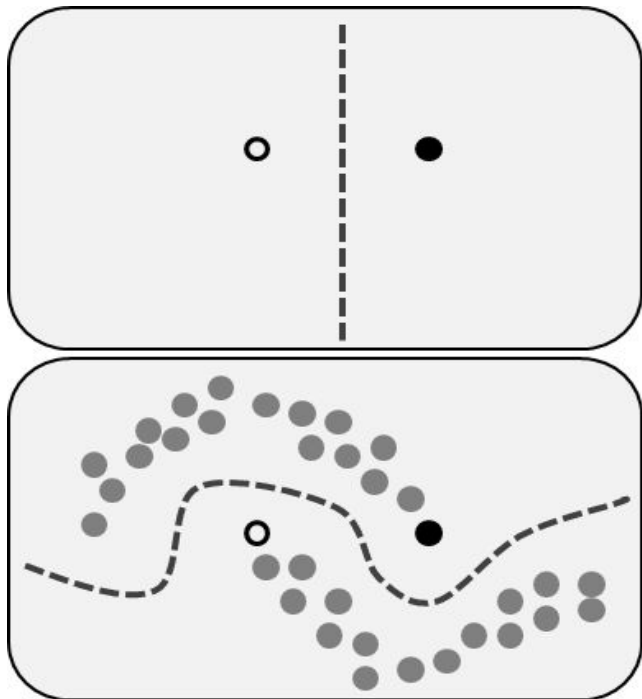


# Discussion: Techniques for Time Domain Astronomy

David C. Stenning  
Postdoctoral Fellow  
SAMSI & Duke University

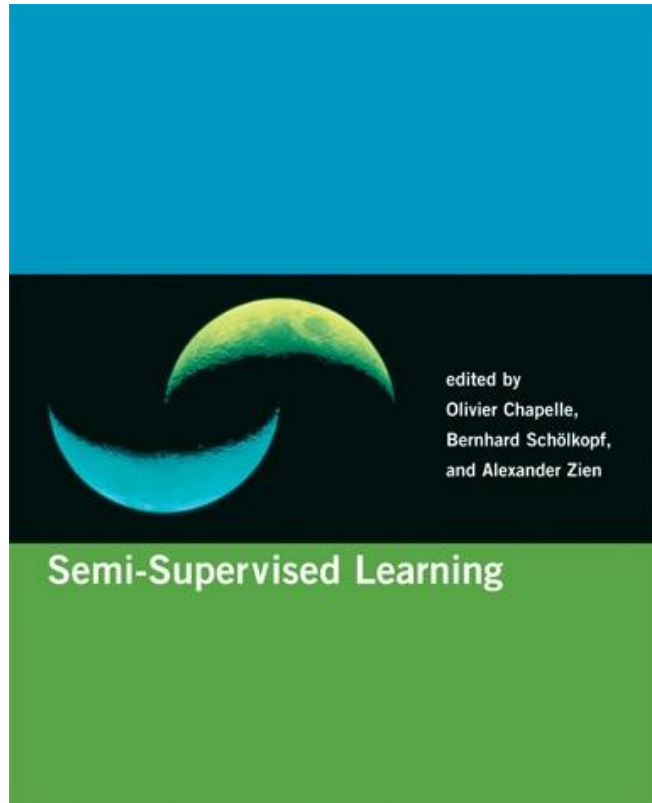
# Semi-supervised learning



- Supervised learning + unlabeled data
- Labeled data may be costly/rare, while unlabeled data cheap/plentiful
- Combine w/ ideas from active learning

Image: [https://en.wikipedia.org/wiki/Semi-supervised\\_learning](https://en.wikipedia.org/wiki/Semi-supervised_learning)

# Semi-supervised learning: (some) references



## Semi-supervised learning for photometric supernova classification<sup>†</sup>

Joseph W. Richards [✉](#), Darren Homrighausen, Peter E. Freeman, Chad M. Schafer, Dovi Poznanski

First published: 28 October 2011 [Full publication history](#)

DOI: [10.1111/j.1365-2966.2011.19768.x](https://doi.org/10.1111/j.1365-2966.2011.19768.x) [View/save citation](#)



# Hierarchy-based prior → improved classification?

Bayesian Analysis (2007)

2, Number 1, pp. 221–238

## Improving Classification When a Class Hierarchy is Available Using a Hierarchy-Based Prior

Babak Shahbaba\* and Radford M. Neal†

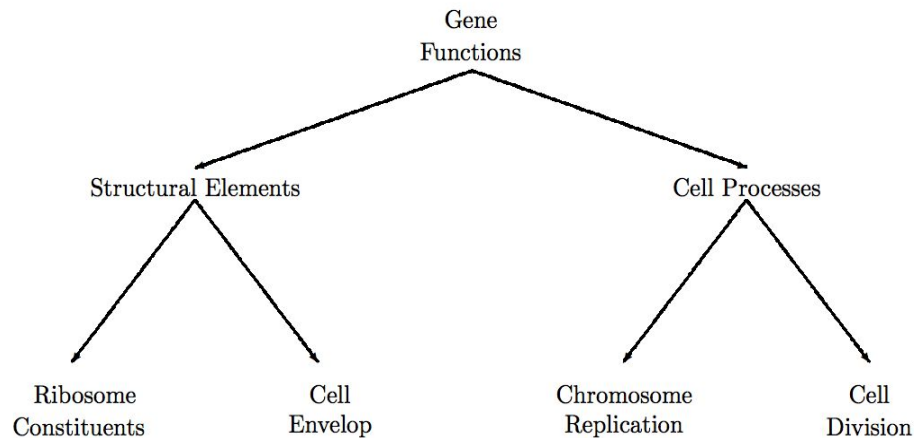


Figure 1: A part of a gene annotation hierarchy proposed by Riley (1993) for the *E. coli* genome.

# Questions for discussion

- Uncertainty quantification (UQ) in a classification context?
- Can we account for uncertainty on “ground truth?” (Should we?)
- Similarly, can we quantify uncertainty on (predicted) labels? And propagate the uncertainty through secondary inferences?
- How should we account for error bars on the features themselves?
- How do we balance the tradeoff between computing features/fitting models and obtaining high classification accuracy?
- Which MCMC (or other) approach is best for a given problem? (Best to start simple and consider more complex methods as needed?)
  - How to treat multiple modes? And what to report (i.e. fitted values and error bars not sufficient w/ multiple modes)?

# Multiple Modes

