

Introduction

My name. My email.

Learning to separate. So learning why and when.

What we'll learn

Things you know and relationships to them:

- Regression (minimise $\| Xb^T - y \|$)
- Logistic regression — think about naive bayesian classifiers — $\Pr(x | y)$ — and then use the (logistic) classifier and learn it directly from the data — $\Pr(y | x)$
- CARTs
- Random forests
- (CARTs and RF are also usable for regression.)

I explain what these things are in high-level terms:

- Logistic Regression
- SVM
- ANN

Note that ANN is not “more advanced” just because it’s second. It’s just harder. And, often, a more efficient way to make time pass without better results.

Format

There will be weekly assignments. We will discuss them in class but I won’t correct them unless you ask.

You will have a final project, which you must do in groups of two (or three in at most one case). The final exam will ask you some questions about your project (individually).

You should probably plan to spend 2–3 hours outside of class for each hour in class. There is, historically, a very strong correlation between your preparation time and your success.

(Discussion based on README here.)

github

There's a github repository.

Recommend use your own laptop if possible. Python and other software is free and easy to install. There's a help page in the git for installing python.

You should create a github repository for the course. *Talk about how to ask questions.* Code questions must be via github issues. (Tag me.) Non-code questions may be by email or github as seems most appropriate, but github issues are pretty convenient.