Introduction

My name. My email.

- Machine learning
- Data science
- Artificial intelligence
- Big data

A better title would be “Fun with Separators”.
“Avoir du plaisir avec des séparateurs”.

What we’ll learn

Things you know and relationships to them:

- Regression (minimise $\parallel Xb^T - y \parallel$) — lasso (minimise $\parallel Xb^T - y \parallel - \lambda \parallel b \parallel$) — ridge (Tikhonov regularisation) (minimise something a bit more complicated)
- Logistic regression — think about naive bayesian classifiers — $\mathbf{Pr}(x \mid y)$ — and then use the (logistic) classifier and learn it directly from the data — $\mathbf{Pr}(y \mid x)$
- CARTs
- Random forests
- (CARTs and RF are also usable for regression.)

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

- SVM — dot product permits kernel trick
- ANN

Note that ANN is not “more advanced” just because it’s second. It’s just harder.

Format

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

There will be a very short quiz at the beginning of each session. It will start and end on time. It will be easy if you did the assignment. If you didn't, it will probably be impossible. (Sometimes it will be writing code.)

(Olivier Darné has asked for the results.)

You will have a final project, which you may do together or individually. There will be a 15 minute oral examination, possibly by video, in which you will be asked some nitty questions to determine that you understand what you did and why.

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.