Removing the blindfold

Visualising statistical models

Hadley Wickham

Assistant Professor Dobelman Family Junior Chair Department of Statistics Rice University



Access









Visualisation	Modelling
 + Uncovers the unexpected - Slow - Cognitive biases 	 Only discovers what was anticipated Fast Mathematically well founded



Neural networks

Using models + visualisation to illuminate how a model works

Display the model in the data space

Look at many members of a collection

Explore the process of fitting, not just the end result

Neural networks

- Modelled on the way that brains work
- Normally treated as a black box. Can we gain more insight into how they work?

• Single hidden-layer neural network: nnet R package



Display the model in data space









How do neural networks work?









Look at all members of the collection









How did I find that model?







Ensembles of linear models

Using models + visualisation to illuminate the underlying data

Display the model in the data space

Look at many members of a collection

Explore the process of fitting, not just the end result

Data

- Fertility in French-speaking Swiss provinces in the late 1800's
- Predict fertility based on:
 - proportion of agricultural workers
 - average performance on an army examination
 - amount of higher education
 - proportion of Catholics
 - infant mortality

Model

- Linear modes with all combinations of covariates (2^p models)
- What can looking at all models tell us that looking at just a few can't?















Conclusions

Other methods

- MANOVA
- Self-organising maps (clusterfly)
- Hierarchical clustering (clusterfly)
- Classification methods (classifly)
- Projection pursuit (tourr)

The future Visualise Model

- Currently iteration between modelling a and visualisation a little clunky: no software has both the visualisation and the modelling capabilities
- We believe R has the most potential, and are working hard to bring interactive graphics into R