Test 2 Info Sheet

Stat 341 – Spring 2017

Logistics

- In-class test: Friday, April 28
- Take-home test: Due Monday, May 1
- Both portions cover Chapters 1-10 of *Statistical Rethinking*, emphasizing material since Chapter 5.
  - Note: We did not cover the portion of Chapter 10 dealing with Poison models
  - Note: Our coverage of Chapter 9 was quite light
  - Note: Multi-level models (aka hiearchical models) will NOT be covered on this test.

Topics

This list isn't meant to be exhaustive, but I hope it is useful in preparing for the test.

Techniques

1. Creating Numerical and Graphical Summaries
   - You have your choice of `lattice`, `ggplot2`, or `ggformula`, but I've been leaning heavily on `ggformula`, since it does what we need most easily. If you use `lattice`, the migration should be pretty easy. If you already know (or are learning) `ggplot2`, it's fine to continue with that. (Anything you can do with `ggformula` can be done with `ggplot2`, since `ggformula` just translates a new interface into `ggplot2` code.)
2. Fitting models with `map()` and `map2stan()`
3. Posterior sampling
   a. `extract.samples()`
   b. `link()`
   c. `sim()`
   d. creating counterfactual data (for use with `link()` and `sim()`, for example)
4. Creating an ensemble model with `ensemble()`
5. Examining models
   a. `precis(model)`, `plot(precis(model))`
   b. `coeftab(model)`, `plot(coeftab(model))`
   c. `compare(...)`, `plot(compare(...))`
   d. `plot(model)` to look at chains of a Stan model
   e. `pairs(model)`
   f. `WAIC(model)`, `DIC(model)`, `logLik(model)`
   g. `show(model)`
6. A non-exhaustive list of R commands you should know
   - Plotting commands (`lattice`, `ggplot2`, or `ggformula` versions) for the types of plots we have been making.
   - `apply()`

- working with data frames: `data_frame()`, `expand.grid()`, `mutate()`, `filter()`, `group_by()`, `summarise()`
- distributions: `dnorm()`, `dbinom()`, `dunif()`, `rnorm()`, `rbinom()`, `runif()`
- miscelaneous: `zscore()`, `str()`, `head()`, `sample()`, `coerce_index()`

Concepts

1. posterior $\propto$ prior $\cdot$ likelihood
2. Overfitting/Underfitting
3. Information Criteria (AIC, DIC, WAIC)
   a. how to calculuate information and entropy
   b. effective number of parameters
   c. Know the connections among `WAIC` and `pWAIC` from the `compare()` output, and `logLik(model)`
4. What posterior sampling is and what can be done with posterior samples
5. Potential problems with quadratic approxiation (`map()`) and why Stan often works better; potential problems with Stan an how to detect them.
6. Metropolis Algorithm
7. Creating models
   - selecting priors that are reasonable
   - creating a model relationship (equation with variables and parameters)
   - linear models, including interaction effects
   - generalized linear models (link functions, etc.)
   - handling categorical variables
8. Interpreting models
   - interpreting model coefficients (and their uncertainty).
   - interpreting model predictions (and their uncertainty).
   - plots that help visualize what a model "thinks"