Week 5: Binomial models

Announcements

1. Sign up for proposal talks
2. For-loops vs. copypasta
3. Parameter recovery & model verification
4. Plotting results
5. Not prior priors

Design matrix activity

Five ways to think about model structure

1. Design matrix
2. R formula syntax
3. Long-form equations
4. Graphical representation
5. Verbal representation

Binomial glm

$$y_i \sim Binom(k_i, p_i)$$

$$\log(\frac{p}{1-p}) = X\beta$$

Why not $p = X\beta$?

Bernoulli glm

Equivalent to binomial with $k = 1$

$$y_i \sim Bernoulli(p_i)$$

$$\log(\frac{p}{1-p}) = X\beta$$

Pro tip:

Logit function: `qlogis()`
Inverse logit function: `plogis()`

Binomial distribution: properties

$y \sim Binom(k, p)$

$$E(y) = kp$$

$$Var(y) = kp(1 - p)$$

Binomial overdispersion

Test with posterior predictive check

2 solutions to overdispersion

1. Binomial-normal model

$$y \sim Binom(k, p)$$

$$\ln\left(\frac{p}{1-p}\right) = X\beta + \epsilon$$

$$\epsilon \sim N(0, \sigma)$$

2. Beta-binomial model

$$y_i \sim Binom(k_i, p_i)$$

$$p_i \sim Beta(\alpha, \beta)$$

Recommendation

1. **Binomial-normal model**

$$y \sim Binom(k, p)$$

$$\ln\left(\frac{p}{1-p}\right) = X\beta + \epsilon$$

$$\epsilon \sim N(0, \sigma)$$

2. Beta-binomial model

$$y_i \sim Binom(k_i, p_i)$$

$$p_i \sim Beta(\alpha, \beta)$$

Caution

Overdispersion is not possible with binary data

Don't try to implement an overdispersed Bernoulli model!

Predictive accuracy

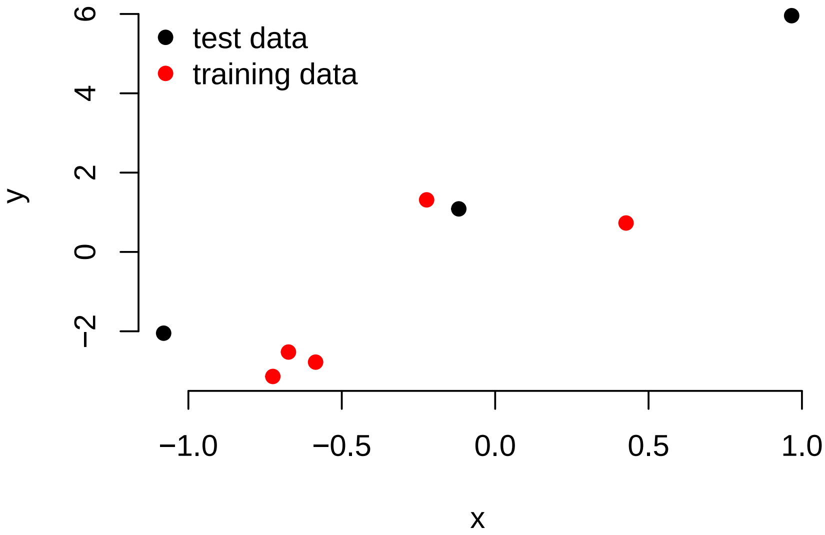
1. Estimate parameters w/ training data:

$\rightarrow [\theta \mid y_{train}]$

2. Make predictions for new observations
3. Compare model predictions to validation data:

- classification error (ROC curves, AUC)
  - good for binary data, but very specific
- validation log likelihood $[y_{test} \mid \theta]$
  - more general
  - easy to compute

Validation log likelihood example



Obtaining estimates with training data

Obtaining estimates with training data

Validation log likelihood

Joint validation log likelihood:

$$\sum_{i=1}^{n_{test}} log([y_{test_i} \mid \theta])$$

Today's class

Mini-Kaggle competition

1. Build a model to classify tumors as malignant or not
2. Evaluate out of sample predictive power
3. Earn prizes