Introduction(week 1, part 3)

6 Jan 2023

Table of contents

books 1
Fisher's irises 1
linear models 2
least squares 2
regression as classification 2
(k)-nearest-neighbor 2
dimensionality 4
bias-variance expansion 4
linear regression 4
references 4

books

- ESL and ADA cover very similar material
- both compare linear regression and nearest-neighbour methods as opposite ends of a specrum

Fisher's irises

- Canadian content (irises of the Gaspé peninsula) (Fisher 1936)
- Fisher was a eugenist (Bodmer et al. 2021)
- multiple versions/errors! (Bezdek et al. 1999)
- alternative: Palmer penguins dataset

Fisher, R. A. 1936. "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics* 7 (2): 179–88. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Bodmer, Walter, R. A. Bailey, Brian Charlesworth, Adam Eyre-Walker, Vernon Farewell, Andrew Mead, and Stephen Senn. 2021. "The Outstanding Scientist, R.A. Fisher: His Views on Eugenics and Race." *Heredity* 126 (4): 565–76. https://doi.org/10.1038/s41437-020-00394-6.

Bezdek, J. C., J. M. Keller, R. Krishnapuram, L. I. Kuncheva, and N. R. Pal. 1999. "Will the Real Iris Data

linear models

- can write out as $\hat{Y} = \hat{\beta}_0 + \sum X_j \hat{\beta}_j$
- go almost immediately to $\hat{Y} = X^\top \hat{\beta}$ or $\langle X, \beta \rangle$ or $\mathbf{X}\beta$
- $\mathbf{X}$ is the *model matrix* (sometimes "design matrix")
- usually includes an intercept column
- can contain *any* (precomputed) functions of input variables
- input vars $\to$ predictor vars
- 1D examples

least squares

- choose L2 norm (p-norm = $\left(\sum |x|^p\right)^{1/p}$)
- poll: why?
- $\sum_i (Y_i - X_i \beta)^2$
- equivalent to $(\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta)$
- differentiate and solve: $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$
- **hat matrix**:

$$\begin{aligned}
\hat{x} = H\mathbf{y} &= \mathbf{X}\hat{\beta} \\
&= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \\
H &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top
\end{aligned}$$

- regression as a *linear filter*

regression as classification

- slightly weird
- looseness of "classification" vs "regression"
  - should probably use *discriminant analysis* here
  - or logistic regression

(k)-nearest-neighbor

- $\frac{1}{k} \sum_{x_i \in N_k(x)} y_i$
- also a linear smoother: columns of $\mathbf{X}$ are $1/k\times$ *indicator variables*: $(x_i \in N_k(x))$ and the β values are y_i

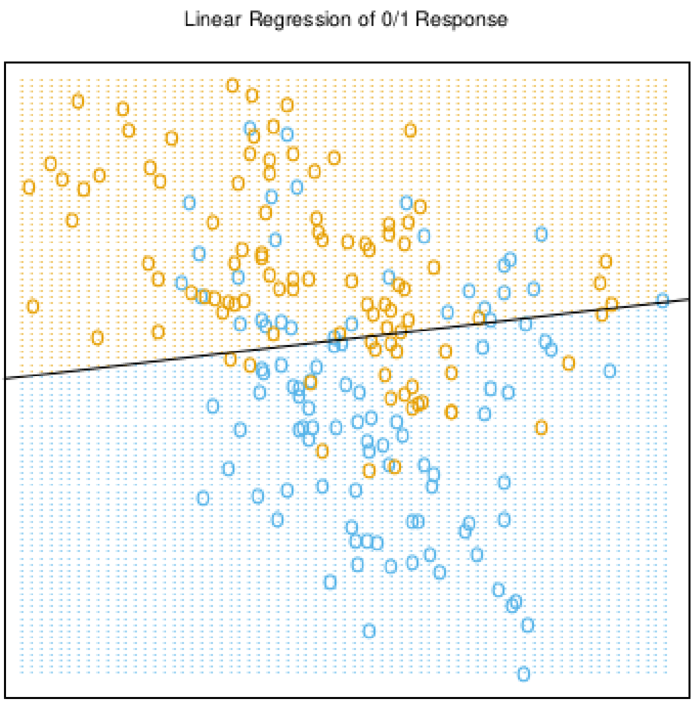

Figure 1: fig2.1

dimensionality

bias-variance expansion

linear regression

references