Lecture 5
GLM computation and data example

Today’s topics:

- GLM computation

- Example: building a GLM

- Reading: Agresti Chapters 4.5, 4.7

GLM computation

- Only discuss the case of $a(\phi) = 1$ to simplify notation
- If $a(\phi)$ is not a constant, one can get $\hat{\beta}$ from the score equations first, and then estimate ϕ from MLE with $\hat{\beta}$ plugged in

Score equation:

$$\dot{L}(\beta) = X^T D V^{-1}(y - \mu) = 0$$

where

$$L(\beta) = \sum [y_i \theta_i - b(\theta_i)] + const.$$

- Newton's method
- Fisher scoring method
- Iteratively reweighted least squares (IRLS): intuitive explanation for Fisher scoring

Newton's method

Second-order approximation of $L(\beta)$

$$L(\beta) \approx L(\beta^{(t)}) + \dot{L}(\beta^{(t)})^T(\beta - \beta^{(t)}) + \frac{1}{2}(\beta - \beta^{(t)})^T \ddot{L}(\beta^{(t)})(\beta - \beta^{(t)})$$

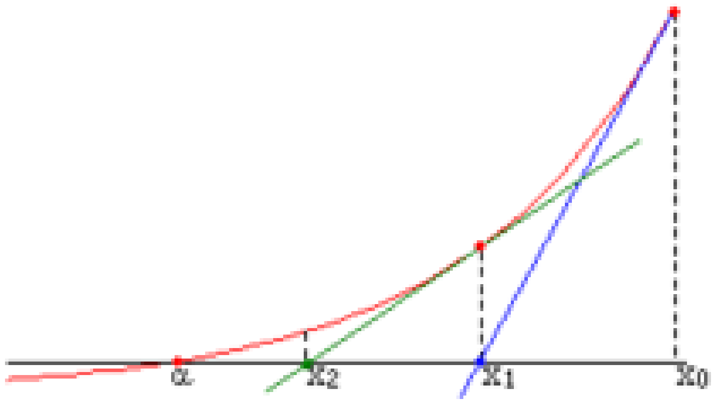
at tth iteration. If $\ddot{L}(\beta^{(t)}) \preceq 0$, then maximizing the second-order approximation is equivalent to solving

$$\dot{L}(\beta) \approx \dot{L}(\beta^{(t)}) + \ddot{L}(\beta^{(t)})(\beta - \beta^{(t)}) = 0$$

We have

$$\beta^{(t+1)} = \beta^{(t)} - \ddot{L}(\beta^{(t)})^{-1}\dot{L}(\beta^{(t)})$$

- Root finding algorithm for solving $\dot{L}(\hat{\beta}) = 0$
  - Local linear approximation of $\dot{L}(\hat{\beta})$

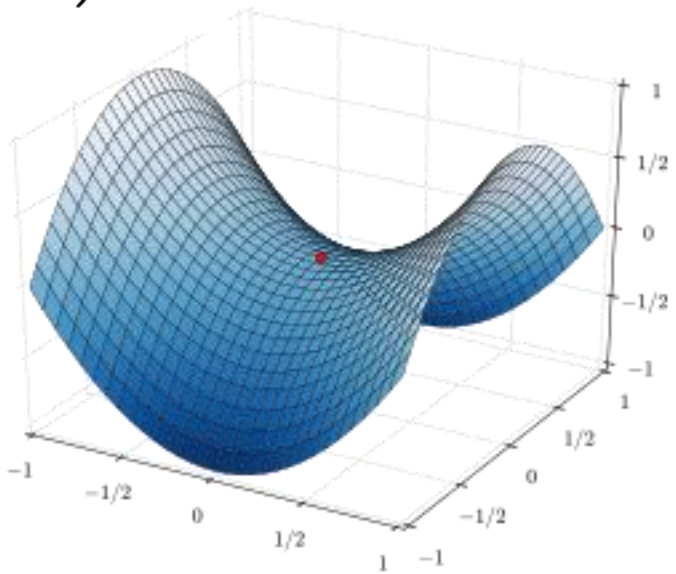


Newton's method

- Newton's method is a general algorithm for optimizing twice-differentiable functions.
- Generally, it converges to the global maximum if $L(\beta)$ is strongly concave
  - If $g(\cdot)$ is the canonical link, then $L(\beta)$ is concave in β

$$-\ddot{L}(\beta^{(t)}) = X^T W^{(t)} X = \frac{1}{a(\phi)^2} X^T V^{(t)} X = -\mathbb{E}\left(\ddot{L}(\beta^{(t)})\right) \succeq 0$$

  - If $g(\cdot)$ is a general link, then $L(\beta)$ is NOT guaranteed to be concave in β
  - If $-\ddot{L}(\beta^{(t)})$ is not non-negative, then step t does not maximize the quadratic approximation (may find a saddle point) and Newton's method may be unstable.



Fisher scoring method

- In lecture 2, we showed that $-\mathbb{E}(\ddot{L}(\beta)) \succcurlyeq 0$ for any β.
- Instead of using the Hessian $\ddot{L}(\beta^{(t)})$ itself in the second order approximation, we use its expectation

$$J^{(t)} = \mathbb{E}\left(\ddot{L}(\beta^{(t)})\right) = -X^T W^{(t)} X$$

Each iteration becomes:

$$\beta^{(t+1)} = \beta^{(t)} - \left(J^{(t)}\right)^{-1} \dot{L}(\beta^{(t)})$$

- For the canonical link, Fisher scoring = Newton's method
- For a general link, Fisher scoring works better in practice

Iteratively reweighted least squares (IRLS)

- We can make a connection between the optimization for GLM and weighted least squares estimation.
- Think of GLM approximately fitting the linear model with transformation on outcome:
$$g(y_i) \sim X_i^T \beta + e_i$$
  - $g(y_i)$ may not be computable
  - e_i should have different variances
  - Assume that after step t, we already have an estimate of $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)$ as $\boldsymbol{\mu}^{(t)} = \left(\mu_1^{(t)}, \cdots, \mu_n^{(t)}\right)$
  - Perform Taylor expansion of $g(y_i)$ at $\mu_i^{(t)}$:
$$g(y_i) \approx g\left(\mu_i^{(t)}\right) + g'\left(\mu_i^{(t)}\right)\left(y_i - \mu_i^{(t)}\right) = X_i^T \beta^{(t)} + g'\left(\mu_i^{(t)}\right)\left(y_i - \mu_i^{(t)}\right)$$
  - Define a "temporary response": $Z_i^{(t)} = X_i^T \beta^{(t)} + g'\left(\mu_i^{(t)}\right)\left(y_i - \mu_i^{(t)}\right)$
  - Then $\mathrm{Var}\left[Z_i^{(t)}\right] = g'\left(\mu_i^{(t)}\right)^2 \mathrm{Var}[y_i] \approx \frac{V_{ii}^{(t)}}{\left(D_{ii}^{(t)}\right)^2} = \left(W_{ii}^{(t)}\right)^{-1}$
  - Fit the linear regression $Z_i^{(t)} \sim X_i^T \beta + e_i$ with weighted least square

Iteratively reweighted least squares (IRLS)

- At the t+1 th iteration, we solve the weighted least square

$$X^T W^{(t)}(z^{(t)} - X\beta) = 0$$

which can be considered as a weighted linear regression with observations $z_i^{(t)}$ and weight w_i for each sample i.

- IRLS is equivalent to Fisher scoring. The tth step of Fisher scoring satisfy

$$\begin{aligned}(X^T W^{(t)} X)\beta^{(t+1)} &= X^T W^{(t)} X\beta^{(t)} + X^T D^{(t)} (V^{(t)})^{-1}(y - \mu^{(t)}) \\ &= X^T W^{(t)} \left[X\beta^{(t)} + (D^{(t)})^{-1}(y - \mu^{(t)})\right] \\ &= X^T W^{(t)} z^{(t)}\end{aligned}$$

- Weight matrix $W^{(t)} \approx \mathrm{Var}\left(z^{(t)}\right)^{-1}$

Example: Building a GLM

- Check Example2 R notebook