STAT 151A Lecture 36

Henry Liev

19 November 2025

Remark 0.1 (Logistic Regression: Final Thoughts)

How to solve for $\hat{\beta}$ numerically?
Goal: want $\hat{\beta}_n$ that maximizes $l(\beta)$, equivalently want $\hat{\beta}_n$ such that
$\frac{\partial l(\hat{\beta}_n)}{\partial \beta} = 0$
Strategy: Newton-Raphson Method:

(1) Start with a guess for $\hat{\beta}_n \to$ call it $\beta_{(0)}$

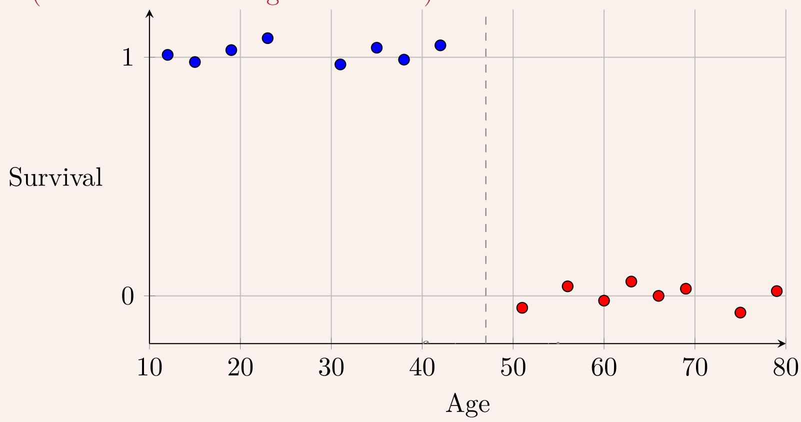
(2) Find tangent line to $l'(\beta)$ at $\beta_{(0)}$

(3) Find the value $\beta_{(1)}$ where tangent line intersects 0

(4) Repeat until convergence

Another way to think about this: repeated Taylor expansions
Why not do gradient descent?
Slower $\to$ main advantage is that you don't need a second derivative $\to$ tangent line requires it
Since 2nd derivative is easy to do here, we use Newton-Raphson
In math: tangent line for step 2:
$l'(\beta_n) = l'(\beta_{(0)}) + (\hat{\beta}_n - \beta_0)l''(\beta_{(0)})$+tiny error
Being more picky about vectors vs scalars, write as:
$\nabla l(\hat{\beta}_n) \approx \nabla l(\beta_{(0)}) - (\hat{\beta}_n - \hat{\beta}_{(0)})I(\beta_{(0)})$
Rearrange in terms of $\hat{\beta}_n$
$\hat{\beta}_n \approx \beta_{(0)} + \left[I_n(\beta_{(0)})\right]^{-1} \nabla l(\beta_{(0)}) \Rightarrow$ set $\beta_{(1)}$ equal to this and iterate until $\beta_{(k+1)} \approx \beta_{(k)}$ ie $\left[I_n(\beta_{(k)})\right]^{-1} \nabla l(\beta_{(k)}) \approx 0$
Plug in $I_n(\beta_{(k)}) = \left(\mathbf{X}^\intercal \mathbf{V}_{(k)} \mathbf{X}\right), \mathbf{V} = \text{diag}(\pi_i(1-\pi_i))$
$\nabla l(\beta_{(k)}) = \mathbf{X}^\intercal y - \mathbf{X}^\intercal \hat{\pi}_{(k)}$
So at convergence, we've solved $\left(\mathbf{X}^\intercal \mathbf{V}_{(k)} \mathbf{X}\right)^{-1} \left[\mathbf{X}^\intercal y - \mathbf{X}^\intercal \hat{\pi}_{(k)}\right] \approx \vec{0}$
Iterable weighted least squares
Our update: $\hat{\beta}_{(1)} = \hat{\beta}_{(0)} + \left(\mathbf{X}^\intercal \mathbf{V} \mathbf{X}\right)^{-1} \left[\mathbf{X}^\intercal y - \mathbf{X}^\intercal \hat{\pi}\right]$
Let $y^* = \mathbf{X}\hat{\beta} + \mathbf{V}^{-1}(y - \hat{\pi})$, then it turns out $\hat{\beta}_{(1)} = \left(\mathbf{X}^\intercal \mathbf{V} \mathbf{X}\right) \mathbf{X}^\intercal \mathbf{V} y^*$

Example 0.2 (Possible dataset: Age vs survival)



Logistic regression doesn't usually allow this $\hat{\pi} = \frac{1}{1+\exp(-\mathbf{X}\hat{\beta})}$
$\hat{\beta}$ will not converge will try to set it infinitely large $\to$ R: "Warning: fitted probabilities of 0 and 1 occurred"
Also "separable" and will lead $\hat{\beta}$ to diverge

Remark 0.3 (Residuals for Logistic Regression)
Recall residual deviance: a bit like ErrSS
$D = -2\sum[y_i \log\hat{\pi}_i + (1-y_i)\log(1-\hat{\pi}_i)]$ comes from log-likelihood
Different versions:

(1) Deviance residuals: $d_i = \pm\sqrt{-2(y_i\log\hat{\pi}_i + (1-y_i)\log(1-\hat{\pi}_i))}, \pm$ determined $\operatorname{sgn}(y_i - \hat{\pi}_i)$

(2) Response residual: $y_i - \hat{\pi}_i$

(3) Can also standardize either deviance or response residuals

  (a) Pearson Residuals: $\frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$

  (b) Standardized Pearson Residuals: $\frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} \cdot \frac{1}{\sqrt{1-h_{ii}}}, \mathbb{H} = \mathbf{X}\mathbf{V}^{1/2}\left(\mathbf{X}^\intercal\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{V}^{1/2}\mathbf{X}^\intercal$

  (c) Standardized Deviance Residual: $\frac{d_i}{\sqrt{1-h_{ii}}}$