STAT 302: Logistic Regression

Sheridan Grant

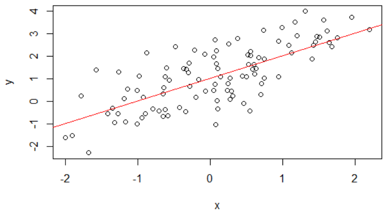
University of Washington Statistics Department

slgstats@uw.edu

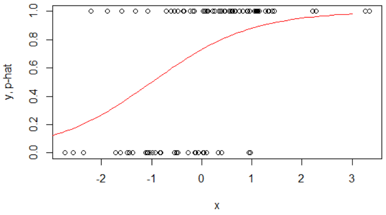
May 13, 2020

Linear vs. Logistic Regression

Linear Regression



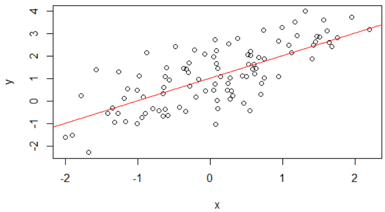
Logistic Regression



Linear vs. Logistic Regression

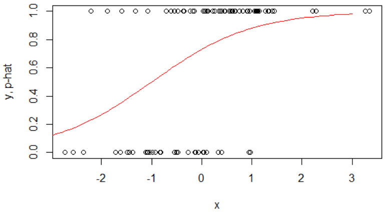
Linear Regression

- Continuous response



Logistic Regression

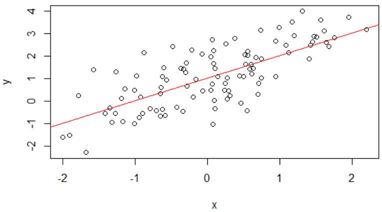
- Binary response



Linear vs. Logistic Regression

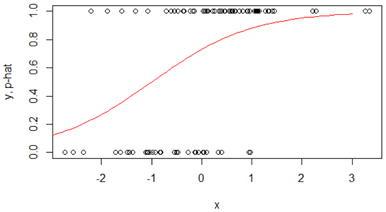
Linear Regression

- Continuous response
- $\hat{Y}_i$ is the predicted outcome for response variable Y



Logistic Regression

- Binary response
- $\hat{p}_i = \hat{P}(Y_i = 1)$ is estimated probability response $Y = 1$



Linear vs. Logistic Regression

Linear Regression

- $\hat{Y}_i$ is a linear function of parameters β

Logistic Regression

- $\text{logit}(\hat{p}_i)$ is a linear function of parameters β

Linear vs. Logistic Regression

Linear Regression

- $\hat{Y}_i$ is a linear function of parameters β
- Estimate model parameters/coefficients by minimizing residuals (you did this yourself on HW 4 using `optim`)

Logistic Regression

- $\text{logit}(\hat{p}_i)$ is a linear function of parameters β
- Estimate model parameters/coefficients by minimizing *deviance* (you don't know this, but R will do it for you!)

Linear vs. Logistic Regression

Linear Regression

- $\hat{Y}_i$ is a linear function of parameters β
- Estimate model parameters/coefficients by minimizing residuals (you did this yourself on HW 4 using `optim`)
- Predictions are $\hat{Y}_i$, can be evaluated on test data Y_i with *RMSE*, *MAE*, etc.

Logistic Regression

- $\text{logit}(\hat{p}_i)$ is a linear function of parameters β
- Estimate model parameters/coefficients by minimizing *deviance* (you don't know this, but R will do it for you!)
- Predictions $\hat{Y}_i$ are a *function* of $\hat{p}_i$, can be evaluated on test data Y_i

A *Generalized* Linear Model

- If Y continuous on $\mathbb{R}$, then $Y = \beta_0 + \beta X$ makes sense. But $\text{P}(Y = 1) \in [0, 1]$!
- Trick: transform $p_i = \text{P}(Y_i = 1)$ so that we can use a linear model

A *Generalized* Linear Model

- If Y continuous on $\mathbb{R}$, then $Y = \beta_0 + \beta X$ makes sense. But $\mathrm{P}(Y = 1) \in [0, 1]$!
- Trick: transform $p_i = \mathrm{P}(Y_i = 1)$ so that we can use a linear model

1. $p \in [0, 1]$

A *Generalized* Linear Model

- If Y continuous on $\mathbb{R}$, then $Y = \beta_0 + \beta X$ makes sense. But $P(Y = 1) \in [0, 1]$!
- Trick: transform $p_i = P(Y_i = 1)$ so that we can use a linear model

1. $p \in [0, 1]$
2. $\frac{p}{1-p} \in [0, \infty)$

A *Generalized* Linear Model

- If Y continuous on $\mathbb{R}$, then $Y = \beta_0 + \beta X$ makes sense. But $\mathrm{P}(Y = 1) \in [0, 1]$!
- Trick: transform $p_i = \mathrm{P}(Y_i = 1)$ so that we can use a linear model

1. $p \in [0, 1]$
2. $\frac{p}{1-p} \in [0, \infty)$
3. $\mathrm{logit}(p) \equiv \log\left(\frac{p}{1-p}\right) \in \mathbb{R}$

A *Generalized* Linear Model

- If Y continuous on $\mathbb{R}$, then $Y = \beta_0 + \beta X$ makes sense. But $\mathrm{P}(Y = 1) \in [0, 1]$!
- Trick: transform $p_i = \mathrm{P}(Y_i = 1)$ so that we can use a linear model

1. $p \in [0, 1]$
2. $\frac{p}{1-p} \in [0, \infty)$
3. $\text{logit}(p) \equiv \log\left(\frac{p}{1-p}\right) \in \mathbb{R}$
4. $\text{logit}(p_i) = \beta_0 + \beta X_i + \epsilon_i$, $\text{logit}(\hat{p}_i) = \hat{\beta}_0 + \hat{\beta} X_i$
5. $\text{logit}^{-1} \equiv \text{expit}$ ($\log^{-1} = \exp$, right?) so $\hat{p}_i = \text{expit}(\hat{\beta}_0 + \hat{\beta} X_i)$

Predictive Power of Logistic Regression

Linear Regression

- Predictive power separate from model fit

Logistic Regression

- Predictive power separate from model fit

Predictive Power of Logistic Regression

Linear Regression

- Predictive power separate from model fit
- Predictive metrics can be similar to fit metrics ($RMSE = \sqrt{\frac{1}{n}\sum_i (Y_i - \hat{Y}_i)^2}$, $RSE = \sqrt{\frac{1}{df}\sum_i (Y_i - \hat{Y}_i)^2}$)
- ...or different ($MAE = \frac{1}{n}\sum_i |Y_i - \hat{Y}_i|$)

Logistic Regression

- Predictive power separate from model fit

Predictive Power of Logistic Regression

Linear Regression

- Predictive power separate from model fit
- Predictive metrics can be similar to fit metrics ($RMSE = \sqrt{\frac{1}{n}\sum_i (Y_i - \hat{Y}_i)^2}$, $RSE = \sqrt{\frac{1}{df}\sum_i (Y_i - \hat{Y}_i)^2}$)
- ...or different ($MAE = \frac{1}{n}\sum_i |Y_i - \hat{Y}_i|$)

Logistic Regression

- Predictive power separate from model fit
- Need to get $\hat{Y}_i$ from $\hat{p}_i$

Predictive Power of Logistic Regression

Linear Regression

- Predictive power separate from model fit
- Predictive metrics can be similar to fit metrics ($RMSE = \sqrt{\frac{1}{n}\sum_i (Y_i - \hat{Y}_i)^2}$, $RSE = \sqrt{\frac{1}{df}\sum_i (Y_i - \hat{Y}_i)^2}$)
- ...or different ($MAE = \frac{1}{n}\sum_i |Y_i - \hat{Y}_i|$)

Logistic Regression

- Predictive power separate from model fit
- Need to get $\hat{Y}_i$ from $\hat{p}_i$
- $\hat{Y}_i = \mathbf{1}[\hat{p}_i > 0.5]$ ($\mathbf{1}$ is 1 if condition is met, 0 otherwise)

Predictive Power of Logistic Regression

Linear Regression

- Predictive power separate from model fit
- Predictive metrics can be similar to fit metrics ($RMSE = \sqrt{\frac{1}{n}\sum_i (Y_i - \hat{Y}_i)^2}$, $RSE = \sqrt{\frac{1}{df}\sum_i (Y_i - \hat{Y}_i)^2}$)
- ...or different ($MAE = \frac{1}{n}\sum_i |Y_i - \hat{Y}_i|$)

Logistic Regression

- Predictive power separate from model fit
- Need to get $\hat{Y}_i$ from $\hat{p}_i$
- $\hat{Y}_i = \mathbf{1}[\hat{p}_i > 0.5]$ ($\mathbf{1}$ is 1 if condition is met, 0 otherwise)
- $\hat{Y}_i = \mathbf{1}[\hat{p}_i < 0.1]$ is dumb. But is $\hat{Y}_i = \mathbf{1}[\hat{p}_i > 0.8]$?

Predictive Power of Logistic Regression

Linear Regression

- Predictive power separate from model fit
- Predictive metrics can be similar to fit metrics ($RMSE = \sqrt{\frac{1}{n}\sum_i (Y_i - \hat{Y}_i)^2}$, $RSE = \sqrt{\frac{1}{df}\sum_i (Y_i - \hat{Y}_i)^2}$)
- ...or different ($MAE = \frac{1}{n}\sum_i |Y_i - \hat{Y}_i|$)

Logistic Regression

- Predictive power separate from model fit
- Need to get $\hat{Y}_i$ from $\hat{p}_i$
- $\hat{Y}_i = \mathbf{1}[\hat{p}_i > 0.5]$ ($\mathbf{1}$ is 1 if condition is met, 0 otherwise)
- $\hat{Y}_i = \mathbf{1}[\hat{p}_i < 0.1]$ is dumb. But is $\hat{Y}_i = \mathbf{1}[\hat{p}_i > 0.8]$?
- Find out with prediction metrics ($ACC = \frac{1}{n}\sum_i \mathbf{1}[\hat{Y}_i = Y_i]$)