Logistic Regression

I Model Set Up

Suppose Y_i are iid RVs w/ binary outcomes.

i.e. $Y_i \overset{\perp\!\!\perp}{\sim} \text{Bern}(\pi_i)$

Bernoulli dist w/ prob π_i

$\Leftrightarrow$ Y_i has pmf:

value	prob
0	$1-\pi_i$
1	π_i

Consider the simple Regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{w/ } Y_i = 0, 1$$

Taking the mean gives

$$E(y_i) = \pi_i = \beta_0 + \beta_1 x_i$$

There are some problems with this model...

① the errors are also binary:

if $Y_i = 1$, $\varepsilon_i = 1 - \beta_0 - \beta_1 x_i$

if $Y_i = 0$, $\varepsilon_i = -\beta_0 - \beta_1 x_i$

$\Rightarrow$ ε_i are not Normally distributed!

② $Var(Y_i) = var(\varepsilon_i) = \pi_i(1-\pi_i)$

$$= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i)$$

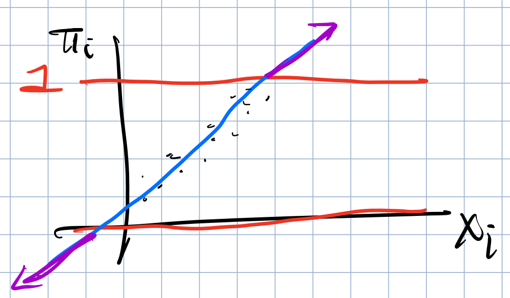
$\Rightarrow$ variance is definitely not constant over X.

$\Rightarrow$ OLS is no longer optimal

③ There are constraints on the response value:

$\pi_i = \beta_0 + \beta_1 X_i$, $\underbrace{0 \leq \pi_i \leq 1}$

In general this constraint isn't met by a linear function:



purple = fitted probs > 1 or < 0.

So modeling π_i as a linear function is probably not what we want to do.

Instead, a nice way to approach this is to use a generalized linear model, where π_i is mapped to some function $g(\pi_i)$ which is unconstrained.

For example, one can take

$$g(t) = \text{logit}(t) = \log\left(\frac{t}{1-t}\right)$$

In this case, the interval $(0,1) \to (-\infty, \infty)$.

We then try to model $g(\pi_i)$ w/ a linear function:

This is called the Logistic regression model.

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i$$

Notice this is a model based on $E(Y_i) = \pi_i$ w/out error term.

Note, if you solve this for π_i, we get:

$$\pi_i = \frac{e^{(\beta_0+\beta_1 X_i)}}{1+e^{(\beta_0+\beta_1 X_i)}}$$

$$= \text{"expit"}(\beta_0+\beta_1 X_i)$$

$\boxed{\text{II}}$ Parameter Estimation

- OLS doesn't work out the box anymore

- Turn to MLE:

$$Y_i \overset{\perp\!\!\!\perp}{\sim} \text{Bern}(\pi_i)$$

$$f(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$\mathcal{L}(\underset{\sim}{\pi}) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$= \prod_{i=1}^{n} \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i}(1-y_i)$$

$\Rightarrow$

$$\ell(\underset{\sim}{\pi}) = \log\mathcal{L}(\underset{\sim}{\pi}) = \sum_{i=1}^{n}\left[y_i \log\left(\pi_i/1-\pi_i\right) + \log(1-\pi_i)\right]$$

Under the logistic regression model:

$$\log\left(\pi_i / 1-\pi_i\right) = \beta_0 + \beta_1 x_i$$

$$\& \quad 1-\pi_i = \frac{1}{1+e^{\beta_0+\beta_1 x_i}}.$$

Plug into $\ell(\underline{\pi})$...

$$\ell(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{n} \log\left(1 + e^{\beta_0+\beta_1 x_i}\right)$$

We want:

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmax}} \, \ell(\beta_0, \beta_1)$$

↳ No nice closed form...

Have to solve using iterative methods.

(See supp. doc.)

Once we have $\hat{\beta}_0$ & $\hat{\beta}_1$, an estimate for π_i is:

$$\hat{\pi}_i = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}$$

Or if we are ok looking at the "odds" instead of absolute probability:

$$\frac{\hat{\pi}_i}{1-\hat{\pi}_i} = e^{\hat{\beta}_0+\hat{\beta}_1 x_i}$$

The odds measures the chance of "success" in a Bernoulli trial relative to the chance of "failure".

Extending to Multiple Logistic Regression...

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_{p-1} x_{(p-1)i}$$

After getting MLEs...

$$\widehat{\text{odds}}_i = \frac{\hat{\pi}_i}{1-\hat{\pi}_i} = e^{\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_{(p-1)} x_{(p-1)i}}$$

III Interpreting Coefficients

For a model w/ only one predictor, for ex:

When $x_i = x$ the odds of success are: $e^{\hat{\beta}_0 + \hat{\beta}_1 x}$

When $x_i = x+1$ the odds of success are: $e^{\hat{\beta}_0 + \hat{\beta}_1 (x+1)}$

The $\boxed{\text{odds ratio}}$ when the predictor x increases by 1 is given by:

$$\frac{e^{\hat{\beta}_0+\hat{\beta}_1(x+1)}}{e^{\hat{\beta}_0+\hat{\beta}_1 x}} = e^{\hat{\beta}_1}$$

Therefore:

- When $\hat{\beta}_1<0$, the odds ratio is <1, & the odds of success are decreasing as x increases.

- When $\hat{\beta}_1>0$, the odds ratio is >1, & the odds of success are increasing as x increases.

- When $\hat{\beta}_1=0$, the odds ratio $=1$, & the odds of success do not change w/ x.

↑ This same story goes for $\hat{\beta}_j$ in MLogR.

What about the "intercept"?

"The odds of success are $e^{\hat{\beta}_0}$ when all predictors have value 0."

<u>Note</u>: The odds of success change by a multiplicative factor of $e^{\hat{\beta}_1}$ or $100(e^{\hat{\beta}_1}-1)\%$ per unit increase in x.

<u>For Ex</u>:

If $\hat{\beta}_1 = 2$, the odds of success increase by $100(e^2-1) = 639.1\%$ for every unit increase in X, on average.

Soft & Hard Prediction

Logistic Regression predicts the chance of $Y_i = 1$: $\hat{\pi}_i$.

In real life, we may want a "hard prediction" — $Y_{new} = 1$ or 0?

To go from our "soft pred" $\hat{\pi}_i$ to a hard prediction, we choose a threshold at π^*:

If $\hat{\pi}_{new} > \pi^*$: predict $\hat{Y}_{new} = 1$

If $\hat{\pi}_{new} \leq \pi^*$: predict $\hat{Y}_{new} = 0$.

The choice of π^* is up to us — a typical default is $\pi^* = 0.5$

In general we can choose π^* such that it optimizes some criteria like

- Accuracy
- Precision
- Recall
- True Positive Rate
- False Positive Rate

more in ML Lab