19.2) Logistic Regression

Vitor Kamada

January 2020

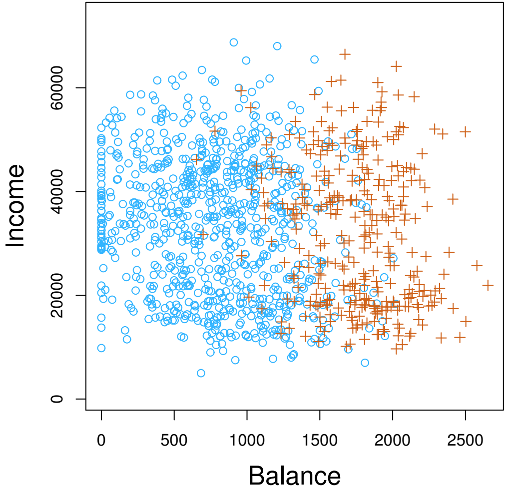
Reference

Tables, Graphics, and Figures from

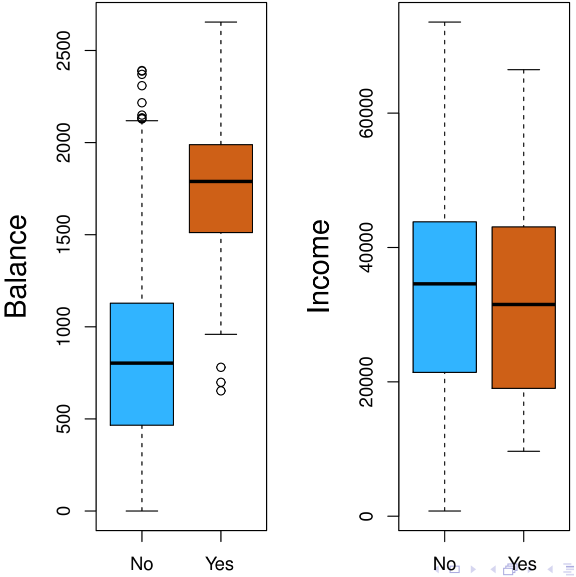
An Introduction to Statistical Learning

James et al. (2017): Chapters: 4.3

The Default Data Set: Default Rate $\cong 3\%$



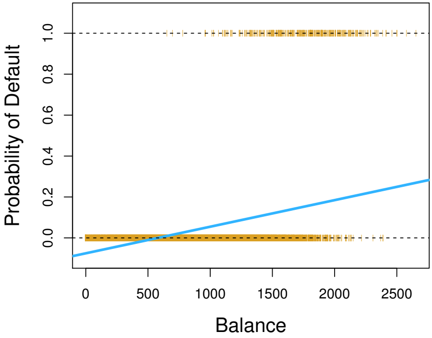
A Subset of 10,000 Individuals

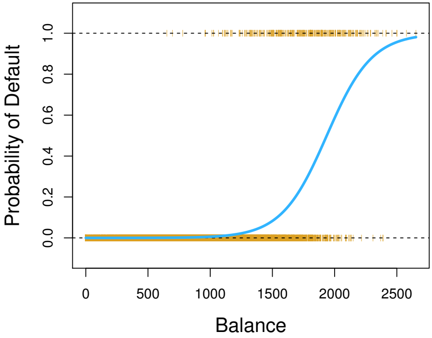


Why Not Linear Regression?

$$y = \begin{cases} 1 \ \ if \ \ stroke \\ 2 \ \ if \ \ drug \ \ overdose \\ 3 \ \ if \ \ seizure \end{cases}$$

Linear Probability vs Logistic Model





The Logistic Model

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

$$1 - p(X) = \frac{1}{1+e^{\beta_0+\beta_1 X}}$$

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$$

$$log\left[\frac{p(X)}{1-p(X)}\right] = \beta_0 + \beta_1 X$$

Logistic Regression

	Coefficient	Std. error	Z-statistic	P-value
Intercept	-10.6513	0.3612	-29.5	<0.0001
balance	0.0055	0.0002	24.9	<0.0001

	Coefficient	Std. error	Z-statistic	P-value
Intercept	-3.5041	0.0707	-49.55	<0.0001
student[Yes]	0.4049	0.1150	3.52	0.0004

	Coefficient	Std. error	Z-statistic	P-value
Intercept	-10.8690	0.4923	-22.08	<0.0001
balance	0.0057	0.0002	24.74	<0.0001
income	0.0030	0.0082	0.37	0.7115
student[Yes]	-0.6468	0.2362	-2.74	0.0062

Predictions given Balance

	Coefficient	Std. error	Z-statistic	P-value
Intercept	−10.6513	0.3612	−29.5	<0.0001
balance	0.0055	0.0002	24.9	<0.0001

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

$$\frac{e^{-10.65 + 0.0055 \times 1{,}000}}{1 + e^{-10.65 + 0.0055 \times 1{,}000}} = 0.57\%$$

$$X = 2000 \rightarrow \hat{p}(X) = 58.6\%$$

Predictions given Student

	Coefficient	Std. error	Z-statistic	P-value
Intercept	−3.5041	0.0707	−49.55	<0.0001
student[Yes]	0.4049	0.1150	3.52	0.0004

$$\hat{Pr}(default = Yes|student = Yes)$$

$$\frac{e^{-3.5+0.405\times 1}}{1+e^{-3.5+0.405\times 1}} = 4.3\%$$

$$\hat{Pr}(default = Yes|student = No)$$

$$\frac{e^{-3.5}}{1+e^{-3.5}} = 2.9\%$$

Multiple Logistic Regression

	Coefficient	Std. error	Z-statistic	P-value
Intercept	−10.8690	0.4923	−22.08	<0.0001
balance	0.0057	0.0002	24.74	<0.0001
income	0.0030	0.0082	0.37	0.7115
student[Yes]	−0.6468	0.2362	−2.74	0.0062

$$\hat{p}(X) = \frac{e^{-10.87+0.0057\times 1{,}500+0.003\times 40-0.65\times 1}}{1+e^{-10.87+0.0057\times 1{,}500+0.003\times 40-0.65\times 1}} = 5.8\%$$

$$\hat{p}(X) = \frac{e^{-10.87+0.0057\times 1{,}500+0.003\times 40-0.65\times 0}}{1+e^{-10.87+0.0057\times 1{,}500+0.003\times 40-0.65\times 0}} = 10.5\%$$