

Chp. 4: Classification

Classification

Predict a qualitative response for an observation based on X .

The 'Default' dataset:

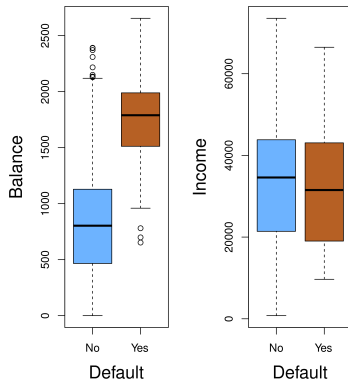


Figure 1: Fig 4.1

Classification

Predict a qualitative response for an observation based on X .

The 'Default' dataset:

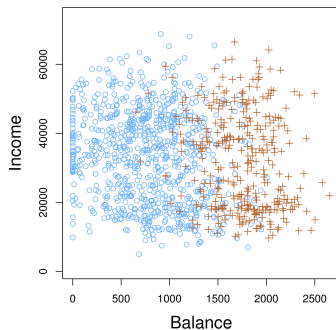
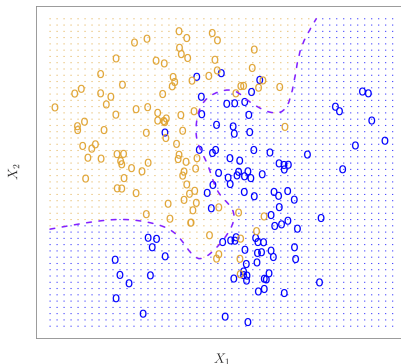


Figure 2: Fig 4.1

Fig. 4.1

Classification

- ▶ The test error rate is minimized on average by a simple classifier (the Bayes classifier) that *assigns each observation to the most likely class, given its predictor values*.
- ▶ With 2 classes, Bayes decision boundary corresponds to predicting class one if $Pr(Y = 1|X = x_0) > 0.5$ and class two otherwise.



Classification

In reality, we never know the conditional probability of Y given X ! The Bayes Classifier boundary is therefore an unattainable 'gold standard'. But, we can estimate it using various methods (logistic regression, KNN classification, LDA, QDA).

Linear regression?

What if we choose

$$Y = \begin{cases} 0 & \text{if Default} = \text{No} \\ 1 & \text{if Default} = \text{Yes} \end{cases}$$

fit a linear regression model, and predict 'Yes' if $\hat{Y} > 0.5$ and 'No' otherwise.

Why not linear regression?

- ▶ No natural ordering of response variables with >2 classes.
- ▶ Estimates fall outside $[0,1]$.

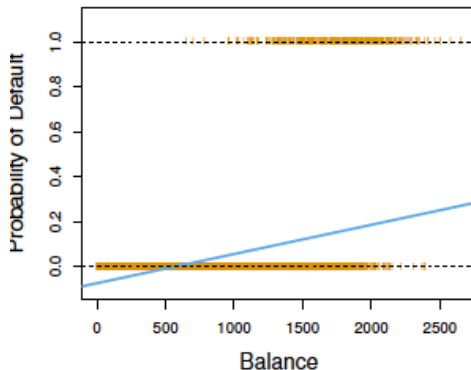


Figure 4: Fig 4.2a

Logistic regression

- ▶ Instead of trying to predict Y , let's try to predict $Pr(Y = 1|X)$, or $p(X)$ for short.
- ▶ Then, predict default = Yes if $p(X) > 0.5$.
- ▶ This threshold can be changed if we want to be more conservative

Logistic regression

- ▶ With logistic regression, we consider the *log-odds* or *logit* transformation of the response, and model the log odds as a linear combination of the predictor variables:

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

Logistic regression

- ▶ Consider the odds, written as: $\frac{p(X)}{1 - p(X)}$
- ▶ e.g. if 1 in 5 people default, the probability of defaulting is 0.2.
- ▶ The odds of defaulting will be 1:4, since $p(X) = 0.2$ and
$$\frac{0.2}{1 - 0.2} = 1/4$$

Logistic regression

By using the logit transformation, our function $p(X)$ takes on an S-shape:

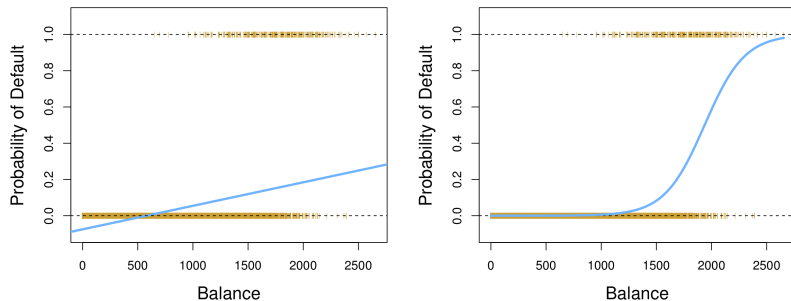


Figure 5: Fig 4.2

Fig. 4.2

Why this results in an S-shaped curve?

Start with the model that we fit for logistic regression:

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

Exponentiate and take multiplicative inverse of both sides:

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

Partial out fraction and add 1 to each side:

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

Why this results in an *S*-shaped curve?

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

Change 1 to a common denominator:

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

Take multiplicative inverse of both sides:

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

Why this results in an *S*-shaped curve?

This is the form of the logistic function:

$$S(x) = \frac{e^x}{1 + e^x}$$

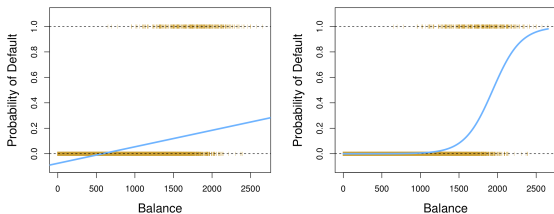


Figure 6: Fig 4.2

Estimating coefficients

- ▶ With logistic regression, we consider the *log-odds* or *logit* transformation of the response, and model the log odds as a linear combination of the predictor variables:

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

- ▶ How did we estimate β_0 and β_1 for linear regression (Chp. 3)?

Estimating coefficients

- ▶ For logistic regression, we use *maximum likelihood* to choose β_0 and β_1 such that the predicted probability $\hat{p}(x_i)$ of the response for each individual is as close as possible to the observed response.
- ▶ β_0 and β_1 are chosen to *maximize* the likelihood function (ISL p. 133).

Fitting a logistic regression model

```
library(ISLR)
data(Default)
summary(glm(default ~ balance, data=Default, family = binomial))

##
## Call:
## glm(formula = default ~ balance, family = binomial, data = Default)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2697  -0.1465  -0.0589  -0.0221   3.7589
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.065e+01  3.612e-01  -29.49  <2e-16 ***
## balance      5.499e-03  2.204e-04   24.95  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpreting the coefficients:

β_1 :

- ▶ For every one unit increase in 'balance', the expected change in the log odds is 0.005.
- ▶ For every one unit increase in 'balance', we expect $e^{0.0055} = 1.0055$ or $\sim 0.55\%$ increase in odds of defaulting on a credit card payment.

Great discussion of interpreting coefficients on UCLA Data Analysis Examples website! See [here](#) for website or [here](#) for logistic regression specifically

Making predictions

Once we have estimated the regression coefficients, we can compute the probability of default for a person with a credit card balance of \$2,000:

$$\hat{p}(X) = \frac{e^{-10.651+0.0055 \times 2,000}}{1 + e^{-10.651+0.0055 \times 2,000}} = 0.586$$

Fitting a logistic regression model

- ▶ Just as for linear regression, we can generalize logistic regression to the case of multiple predictors.

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

NOTE: $p(X)$, the probability of Y given X , is not the same as p , the number of predictors!!!

Fitting a logistic regression model

```
##
## Call:
## glm(formula = default ~ balance + student, family = binom
##       data = Default)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4578  -0.1422  -0.0559  -0.0203   3.7435
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.075e+01  3.692e-01 -29.116  < 2e-16 ***
## balance      5.738e-03  2.318e-04  24.750  < 2e-16 ***
## studentYes   -7.149e-01  1.475e-01  -4.846  1.26e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## (Dispersion parameter for binomial family taken to be 1)
```

Fitting a logistic regression model

$$\begin{aligned}\widehat{Pr}(\text{default}=\text{Yes}|\text{student}=\text{Yes}) &= \frac{e^{-10.75+0.0057 \times 2,000 + -0.71 \times 1}}{1 + e^{-10.75+0.0057 \times 2,000 + -0.71 \times 1}} \\ &= 0.48\end{aligned}$$

$$\begin{aligned}\widehat{Pr}(\text{default}=\text{Yes}|\text{student}=\text{No}) &= \frac{e^{-10.75+0.0057 \times 2,000 + -0.71 \times 0}}{1 + e^{-10.75+0.0057 \times 2,000 + -0.71 \times 0}} \\ &= 0.66\end{aligned}$$

Given the same credit card balance, a student is less likely to default on the payment than a non-student.

A machine learning view on logistic regression

<https://towardsdatascience.com/breaking-it-down-logistic-regression-e5c3f1450bd#6a7a>

Alternatives to logistic regression

- ▶ When classes are well-separated, parameter estimates for logistic regression can be unstable.
- ▶ If n is small and distribution of the predictors is approximately normal in each of the classes, **linear discriminant** model is more stable than logistic regression.
- ▶ Extensions to >2 classes exist (e.g. multinomial logistic regression) but less popular