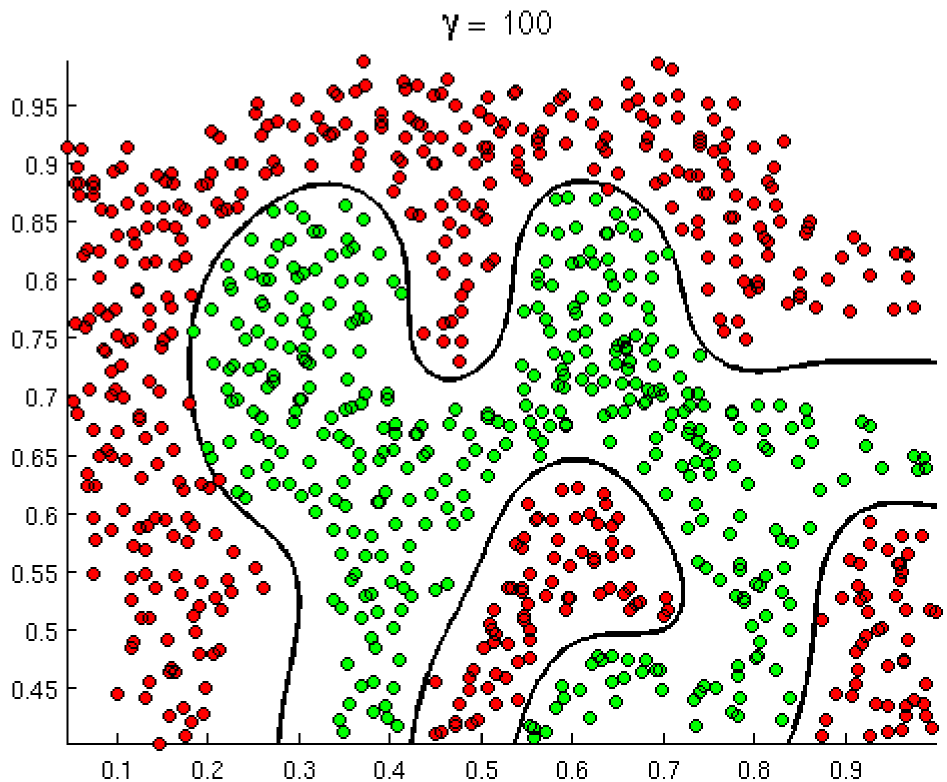
Model Building and Goodness of Fit: (Case Study: Logistic Regression)



STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

Reminder: Logistic Regression

- Logistic Regression:
  - $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_k x_k^{(i)}$
  - π_i: the likelihood that $y^{(i)} = 1$
- (Log-)Likelihood:
  - Compute π_i for every datapoint for which $y^{(i)} = 1$, and $(1 - \pi_i)$ for every datapoint for which $y^{(i)} = 0$.
  - If the fit is very good, the product $\mathrm{P}(\mathrm{y}|\beta) = \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$ is close to 1
    - $\log P(y|\beta) = \sum_{i=1}^{n}(y_i \log \pi_i + (1 - y_i)\log(1 - \pi_i))$ is close to *0*

Deviance

$$deviance = const - 2\log \mathrm{P}(\mathrm{y}|\beta)$$

- (where β is the fitted parameter – the one that maximizes $\log \mathrm{P}(\mathrm{y}|\beta)$. In other words, $\log \mathrm{P}(\mathrm{y}|\beta) = LMAX$)
- Smaller deviance => better fit
  - “Better fit” means π_i is close to 1 if y_i is close to 1, and π_i is close to 0 if y_i is close to 0

Last Time

- Null Hypothesis: the extra coefficients in the full model are 0
- Test Statistic:
  - $LRT = 2\log(LMAX_{full}) - 2\log(LMAX_{reduced})$
  - Has a χ^2 distribution with df=(#of extra parameters in the full model)
- Test?

Last Time

- Null Hypothesis: the extra coefficients in the full model are 0
- Test Statistic:
  - $LRT = \ 2\log(LMAX_{full}) - 2\log(LMAX_{reduced})$
  - Has a χ^2 distribution with df=(#of extra parameters in the full model)
- Test?
  - (1-pchisq(L, df=df))

Which Covariates to Include?

- $AIC = Deviance + 2 \times p$
  - p – number of parameters
  - Popularly called “Akaike Information Criterion
    - Hirutogu Akaike calls it “An Information Criterion”

- Smaller AIC => Better Model
  - Note: larger model always implies smaller deviance
    - Problem: Why?
  - AIC compensates for that
    - The full model has to be reduce the deviance by enough to be considered better than the reduced model

Classification (Iris Example)

Visualization

- (in R)

Classification

- Classification:
  - Given data, we want to predict the class of the datapoint
- Fit a logistic model and pick a cut point
  - Default: 0.5
- If $\hat{\pi}^* > 0.5$, predict $y^* = 1$
- If $\hat{\pi}^* < 0.5$, predict $y^* = 0$

Confusion Matrix

	+	−	
D	TP	FN	Sensitivity = TP / (TP + FN)
D^c	FP	TN	Specificity = TN / (TN + FP)
	PPV = TP / (TP + FP)	NPV = TN / (TN + FN)	

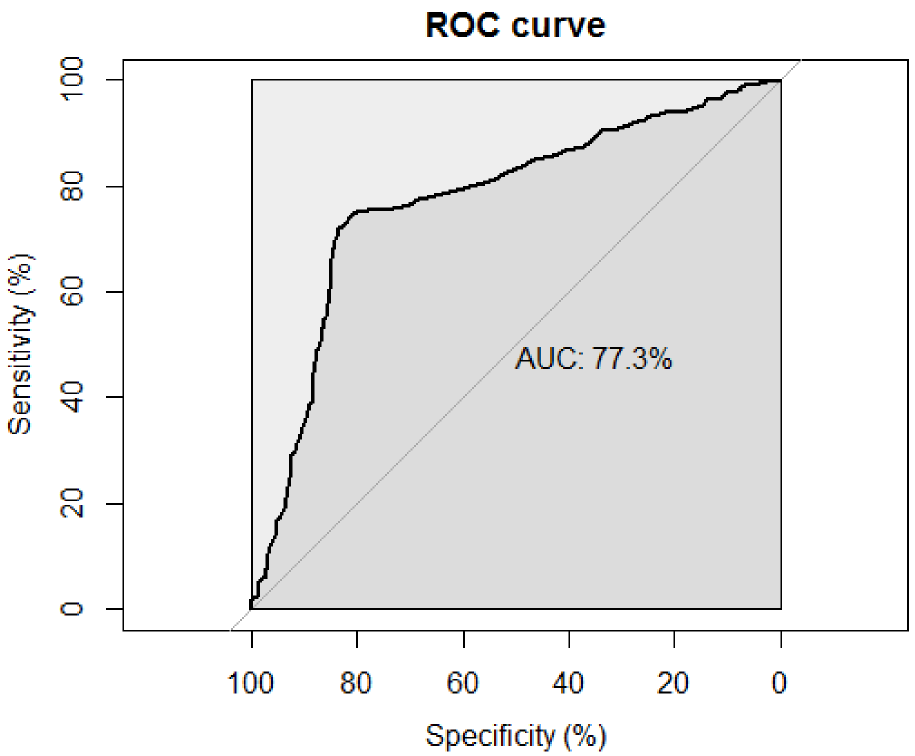
ROC Curves

- Receiver Operating Characteristic
- A plot of sensitivity vs. specificity (complement)
- Originally designed to grade radar detection methods for German planes
- Decades later, their usefulness in classification problems was realized
  - But the name stuck

ROC Curve

```
require(pROC)

> titan.roc <- with(titanR, roc(Isurvived, p, percent=T, auc=T, plot=T,
auc.polygon=T, max.auc.polygon=T, print.auc=T, main= "ROC curve"))
```



AUC

- Area Under the Curve
  - Larger is better
  - Ideally: 100% sensitivity for any specificity