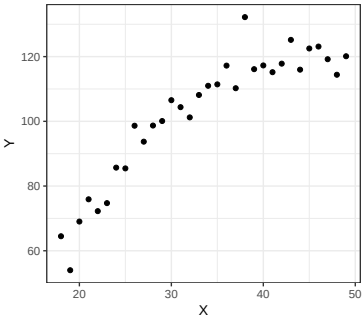
Foundations of Statistical and Machine Learning for Actuaries
Classification, Logistic Regression and Trees

Edward (Jed) Frees, University of Wisconsin - Madison
Andres M. Villegas, University of New South Wales

July 2025

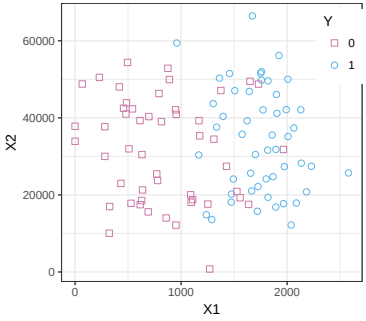
Regression vs. classification

Regression



- Y is quantitative, continuous
- Examples: Sales prediction, claim size prediction, stock price modelling

Classification



- Y is qualitative, discrete
- Examples: Fraud detection, face recognition, accident occurrence, death

Classification problems

- Coding in the binary case is simple:

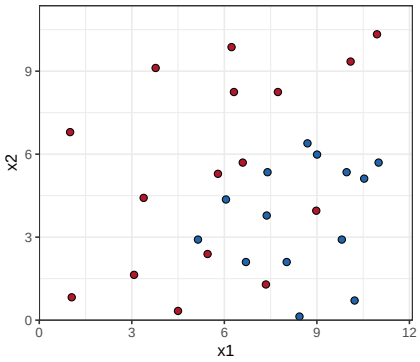
$$Y \in \{0, 1\} \Leftrightarrow Y \in \{\bullet, \bullet\}$$

- Our objective is to find a good predictive model f that can:
  1. Estimate the probability $\Pr(Y = 1|X) \in \{0, 1\}$

$$f(X) \rightarrow \bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet$$

  2. Classify observation

$$f(X) \rightarrow \hat{Y} \in \{\bullet, \bullet\}$$



Can we predict if a road accident will be fatal?

Output (Y):

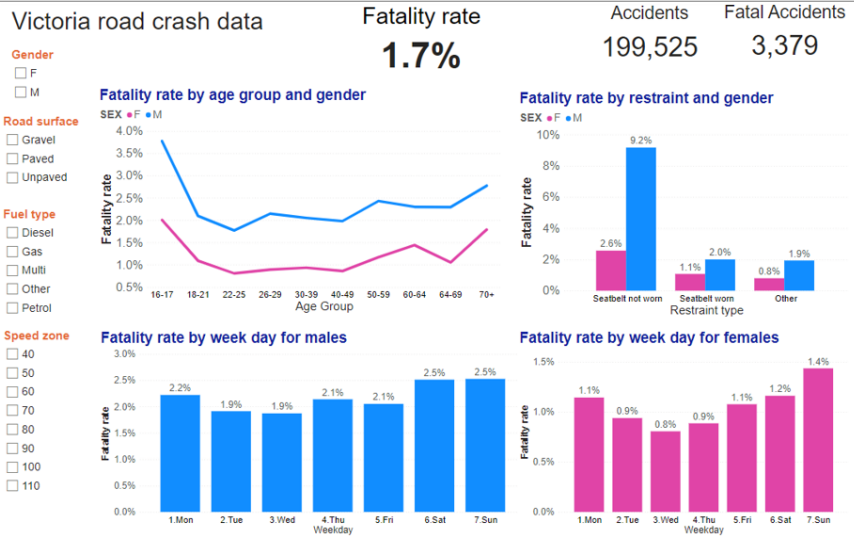
- The accident is fatal; the accident is not fatal

Input (X):

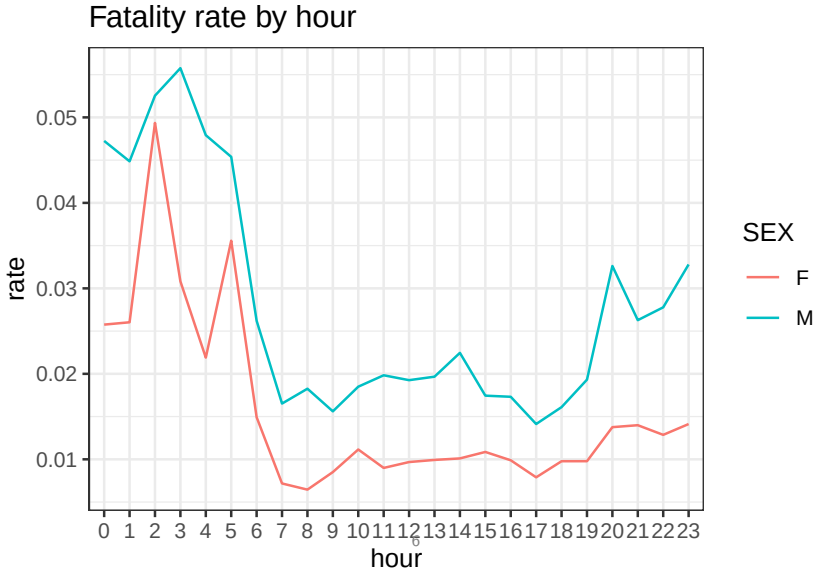
- Age of Driver
- Sex of Driver
- Time of the accident
- Weather conditions
- Type of vehicle
- ...

Source: https://discover.data.vic.gov.au/dataset/crash-stats-data-extract

VicRoads Crash Data



VicRoads Crash Data



Logistic regression

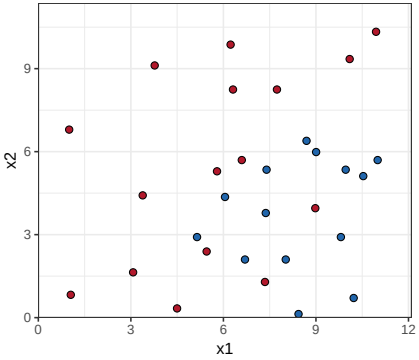
- Perform regression on:

$$\Pr(Y = 1|X) = p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}$$

- In other words:

$$\underbrace{\ln\left(\frac{p(X)}{1-p(X)}\right)}_{\text{log-odds}} = \underbrace{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}_{\text{linear model}}$$

- Use (training) data and maximum-likelihood estimation to produce estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_p$

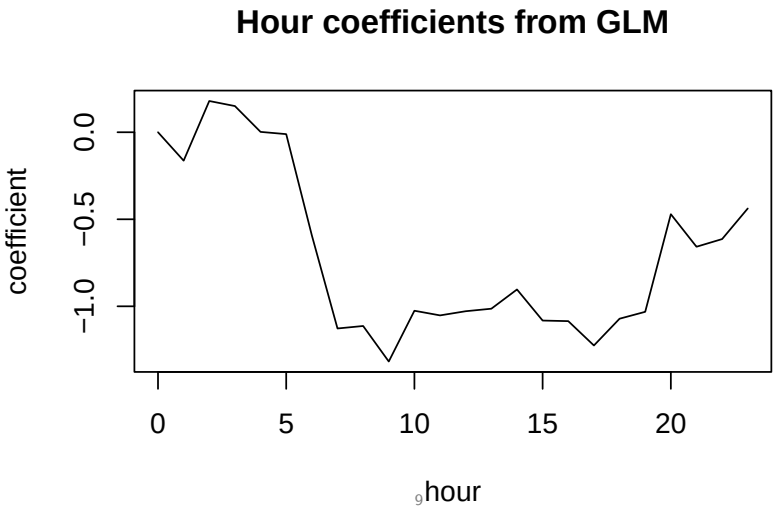


Logistic regression: VicRoads Crash Data

	Estimate	Pr(>|z|)
(Intercept)	-3.49	$< 2e-16^{***}$
SEXM	0.67	$< 2e-16^{***}$
SEXU	-0.38	0.59
HELMET_BELT_WORNSeatbelt not worn	1.50	$< 2e-16^{***}$
HELMET_BELT_WORNSeatbelt worn	0.12	0.01^{*}
AGE_GROUP18-21	-0.31	0.17
AGE_GROUP22-25	-0.50	0.03^{*}
⋮	⋮	⋮
AGE_GROUP70+	0.21	0.35
Weekday2.Tue	-0.13	$0.09^{\cdot}$
Weekday3.Wed	-0.20	0.01^{**}
Weekday4.Thu	-0.06	0.40
Weekday5.Fri	-0.10	0.14
Weekday6.Sat	0.00	1.00
Weekday7.Sun	0.02	0.79

Interpretation: The odds of an accident with a male driver being fatal are $\exp(0.67) = 1.95$ times higher[8] than those of a female driver.

Logistic regression: VicRoads Crash Data



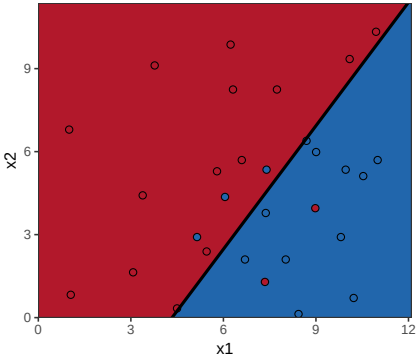
Assessing accuracy in classification problems

- We assess model accuracy using the error rate

$$\text{error rate} = \frac{1}{n}\sum_{i=1}^{n} I(y_i \neq \hat{y}_i)$$

- In our toy example with a 50% threshold

$$\text{training error rate} = \frac{5}{30} = 0.1667$$

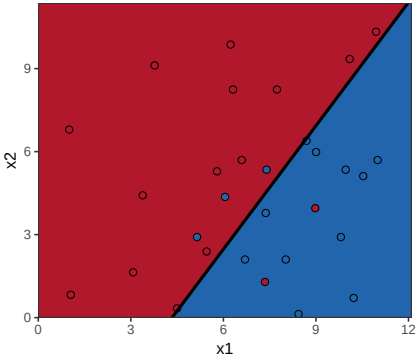


Confusion matrix (50% Threshold)

- Confusion matrix

	$Y = 0$	$Y = 1$	Total
$\hat{Y} = 0$	12	3	15
$\hat{Y} = 1$	2	13	15
Total	14	16	30

- True-Positive Rate $= \frac{13}{16} = 0.875$
- False-Positive Rate $= \frac{2}{14} = 0.1428$

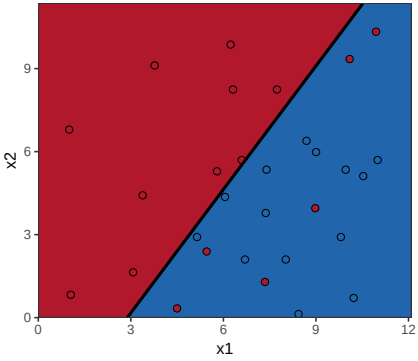


Confusion matrix (25% Threshold)

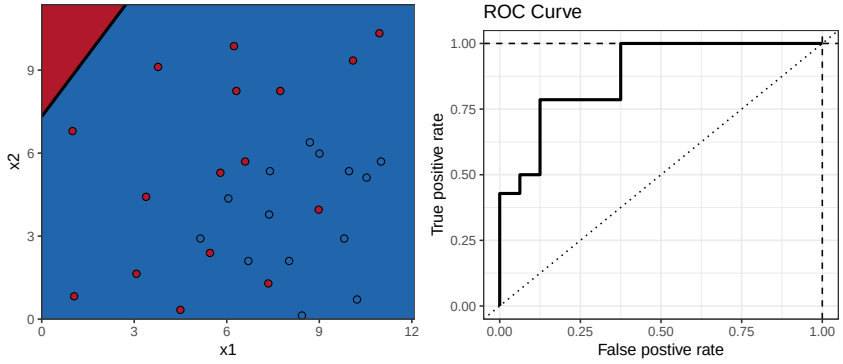
- Confusion matrix

	$Y = 0$	$Y = 1$	Total
$\hat{Y} = 0$	10	0	10
$\hat{Y} = 1$	6	16	22
Total	14	16	30

- True-Positive Rate $= \frac{16}{16} = 1$
- False-Positive Rate $= \frac{6}{14} = 0.4286$

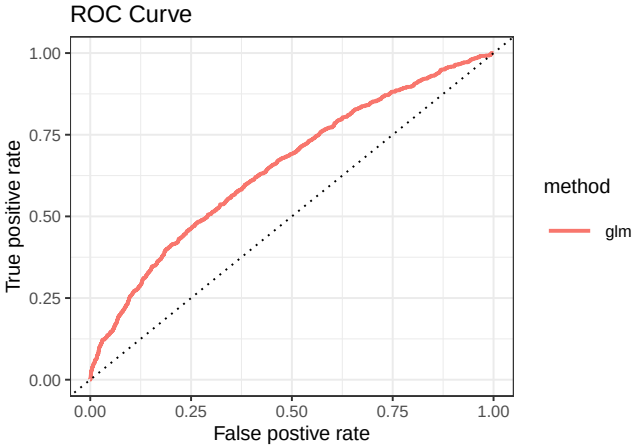


ROC Curve and AUC



- ROC Curve: Plots the true-positive rate against the false-positive rate
- A good model will have its ROC curve hug the top-left corner more
- AUC is the area under the ROC curve: For this toy example AUC=0.8795

Accuracy: VicRoads Crash Data



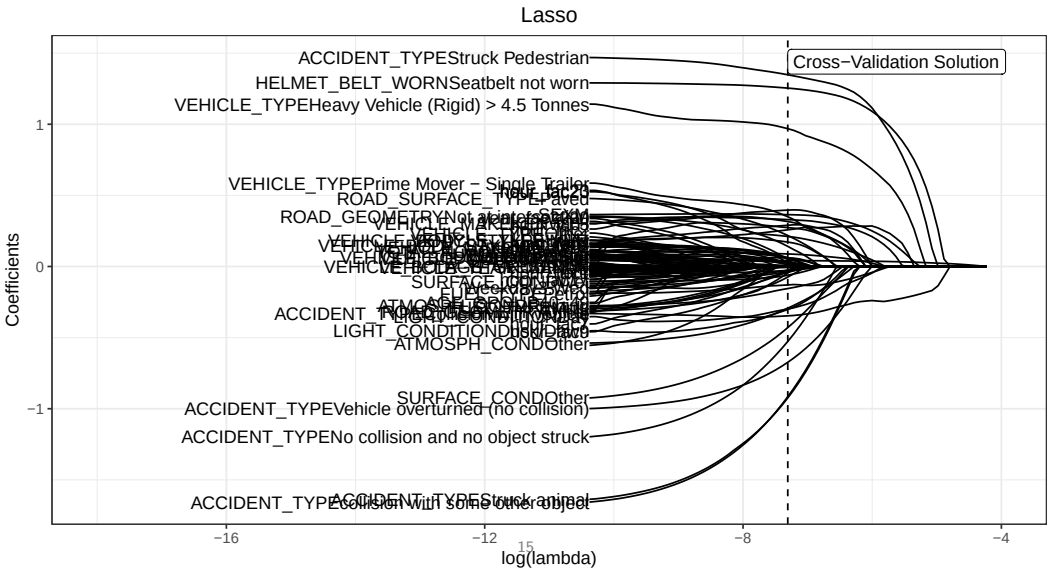
On Test Data

- Confusion matrix

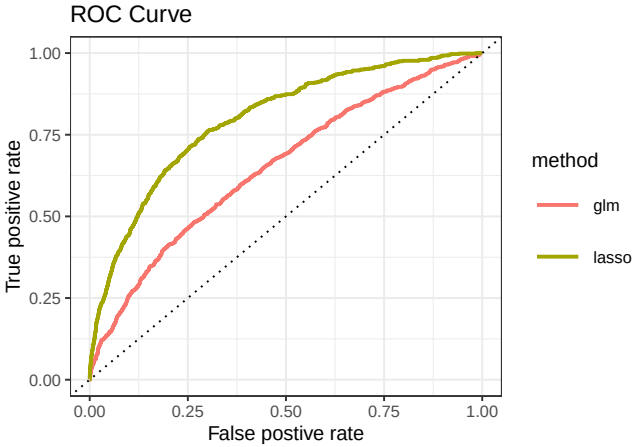
	$Y = 0$	$Y = 1$
$\hat{Y} = 0$	39324	676
$\hat{Y} = 1$	0	0

- Error Rate $= 0.01695$
- Accuracy $= 0.98305$
- AUC$=0.6498$

Logistic + Lasso: VicRoads Crash Data



Logistic + Lasso: VicRoads Crash Data (ROC)



AUC

Method	Train	Test
glm	0.663	0.650
lasso	0.809	0.797

Tree based methods

Tree-based methods

- Stratify / Segment the predictor space into a number of simple regions
- The set of splitting rules can be summarised in a tree

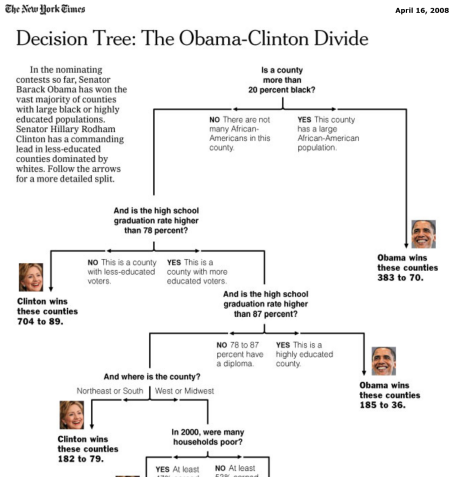
Bagging, random forests, boosting

- Ensemble methods
- Produce multiple trees
- Improve the prediction accuracy of tree-based methods
- Lose some interpretation

Tree based methods: Motivation

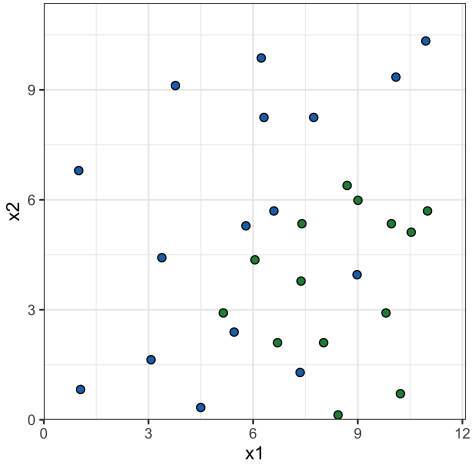
Trees are

- Simple
- Useful for interpretation
- Very common



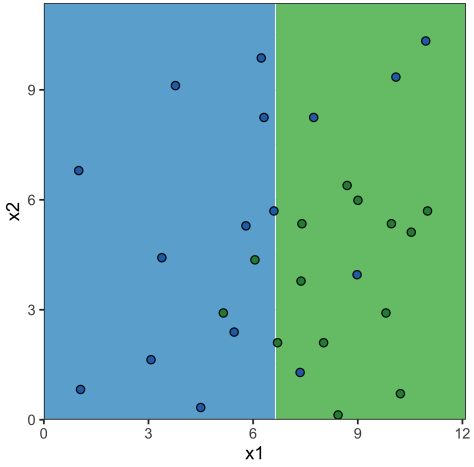
Source: New York Times (2008), Decision Tree: The Obama-Clinton Divide.

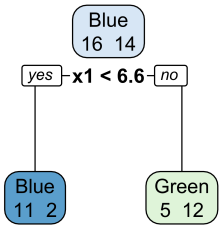
Growing a Tree I



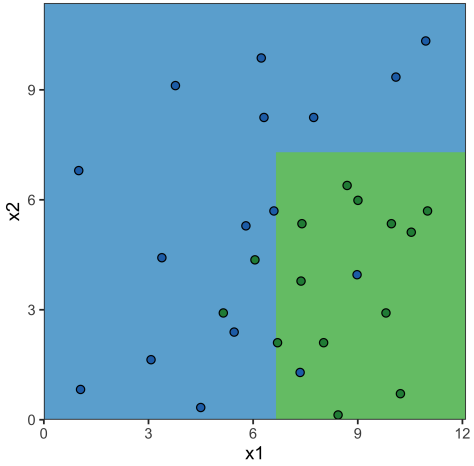
Blue
16 14

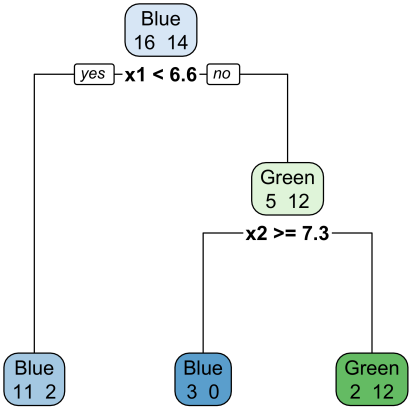
Growing a Tree II



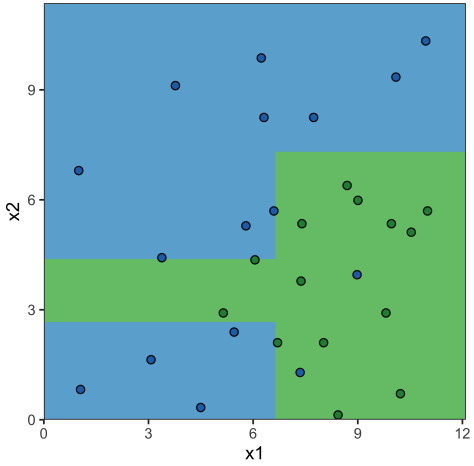


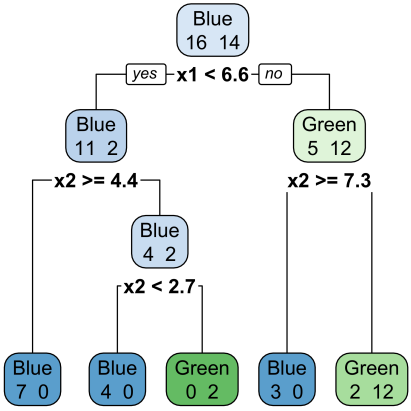
Growing a Tree III



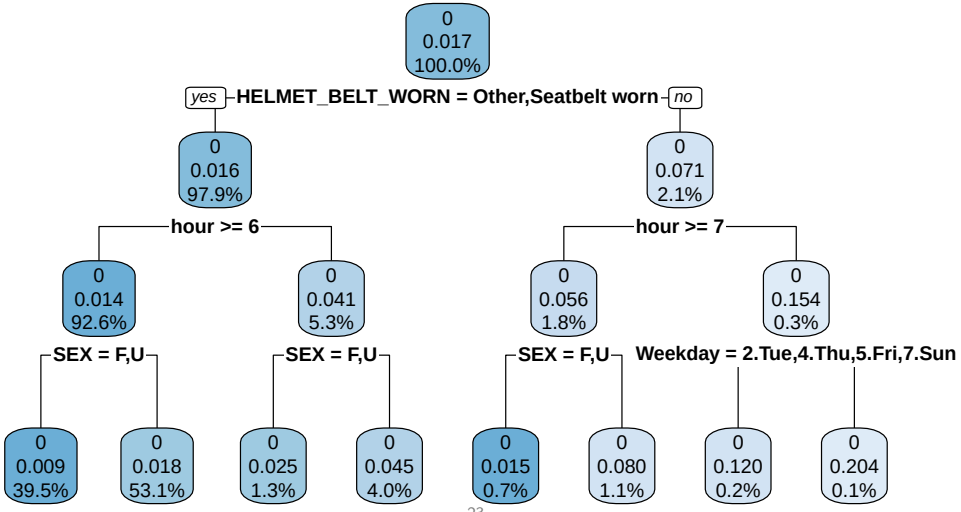


Growing a Tree IV

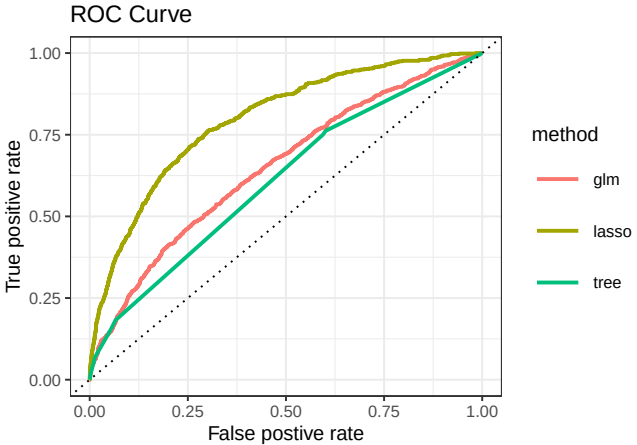




Tree: VicRoads Crash Data



Tree: VicRoads Crash Data (ROC)



AUC

Method	Train	Test
glm	0.663	0.650
lasso	0.809	0.797
tree	0.629	0.610

Advantages and disadvantages of Trees

Advantages

- Easy to explain
- (Mirror human decision making)
- Graphical display
- Easily handle qualitative predictors

Disadvantages

- Low predictive accuracy compared to other regression and classification approaches
- Can be very non-robust

Is there a way to improve the predictive performance of trees?

- Ensemble methods
- Bagging, random forest, boosting

Summary of key concepts in classification problems

We have discussed key concepts in classification problems

- Logistic Regression
- Assessing model accuracy
  - Confusion matrix
  - ROC curve
  - AOC
- Tree-based methods
  - Bagging
  - Random Forest
  - Boosting