Answer 1a

For a fixed x, we can predict c(x) = D or $c(x) = \overline{D}$, for each of these case we can incur 2 types of losses:

$$L(D,D) = 0$$

$$L(\overline{D},\overline{D}) = 0$$

$$L(D,\overline{D}) = 1$$

$$L(\overline{D},D) = 1$$

Answer 1b

The best classification strategy for this case would be the Bayes classifier because we know the conditional probabilities to effectively implement the Bayes classifier. Bayes rule:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

So, using the given formula we will calculate probabilities for both the classes D and \overline{D} and whichever will be the greatest we will assign the variable to that class.

Answer 2a

We have trained model using the training data.

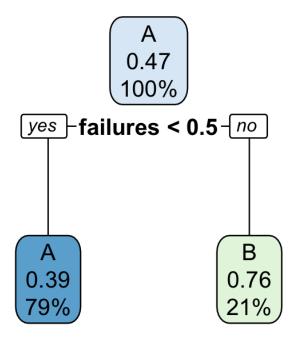
Answer 2b

Confusion Matrix for training data is below:

class_predict_2					
		1	2	3	
	1	29	23	203	
	2	18	21	214	
	3	163	177	1652	

Answer 3a

We have created class A for G3>10 and class B for G3 <= 10. Below is the decision tree created after pruning it from overfitting.



Answer 3b

The generalization error for this decision tree is 0.05158.

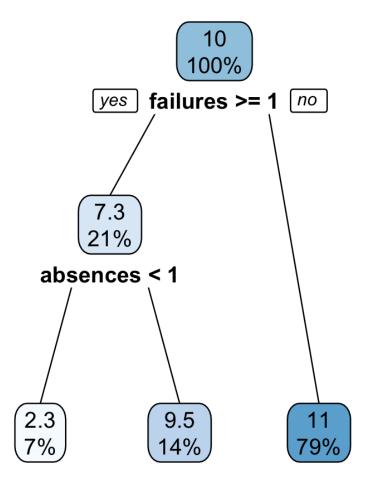
The error rate on training data can be found in the 4th column of the table generated by printcp function in R, which is equal to 0.78495.

Answer 3c

Since the generalization error rate after 1st cut is almost consistent we can say variable failure is the most important feature as we are making division on basis of failures > 0.5.

Answer 3d

Below is the decision tree generated when we used method = "anova".



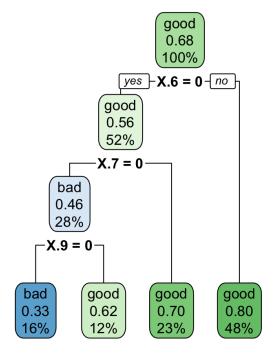
For regression problems, we calculate SSE, and from the predictions of the above regression tree our sum of squared error = 4923.775.

Answer 4

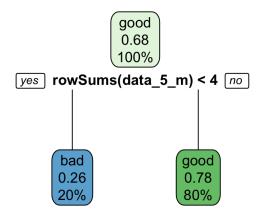
For the first case, we can see that we can separate circles from triangles using a straight line x = y. So, for the new feature if x > y for a particular point it will belong to the class triangle, and vice-versa it will belong to the class circle.

For the second case, we can see circles form a circle on the plane and triangles are outside that circle, from the graph we can assume that circle has a radius = 1and center at (0,0). So equation for this circle becomes, $x^2 + y^2 = 1$. If any point is inside the circle we will predict it as a class circle and for any point outside the circle, we will predict it as a class triangle. The new feature will be $x^2 + y^2 = 1$, then the point belongs to the class circle, else, it belongs to the class triangle.

Answer 5



For this tree, accuracy is 72.5%.



The new feature is the sum of all the features and for this decision tree, the accuracy is 77.5%.

Answer 6a

3					
Ans. 6 a	Suppose we divide Node N mith class				
	propabilities P into Ne 2 Nr				
	neith class probabilities Pe & Pr				
	propabilitées P into Ne 2 Nr nuith class probabilitées Pe 2 Pr 2 proportions que 2 qr				
	Average that enmopy of this division				
	Average that enmopy of this division				
	((eq ((1)) > log (((1)) - (1))				
	90 H(P1) + 9rH(Pr)				
	= 100 (10) = 0 (-100 (P1))+90(-log(P))				
	=> 9 10g (1/2) + 9r wg (1/pr) = 9e(- 20g (Pr))+9r(- 20g(fr))				
	if we the equation process + 9 x(22) / x(pa) (qaz)				
	if we use equation $pk(x_1) + qk(x_2) > k(px_1+qx_2)$ and put $k = -\log p = q_1 > x_2 = p_1 , q = q_1$				
	our equation becomes: -que log Per- que log Pr > - log (que pa + qr Pr)				
	- q log Per - 9 r log Pr > - log (9ePe + 9 r Pr)				
	0 0				
	=> 92 log Pe + 9r log Pr < log (9e Pe + 9r Pr)				
	9, Pe + 9rPr = E(P)				
	2) Quant + Quant (Ing (F(P)) - (1)				
	91 Pe + 9r Pr = E(P) => 91 log Pe + 9r log Pr < log (E(P)) - 0				
	Now consider enmopy at node N $H(P) = E(\log(VP))$ it is given that $E(\log(VP)) \leq \log(E(VP))$				
	$M(P) = C(\log(VP))$				
	it is given that				
	E(LOG (1/p)) { LOG (E(1/p))				

Suppose Twe directle Mode In with close
propositivies P into N, 2 Nr
MICH CLASS DROBORDICHES P. & P.
- E (log (P)) & - log (E(P))
which becomes yours told sooned
ad lim
E (LOG (P)) >, Log (E(P)) - 2
from equation (1) 2 @ we can
dearly say that
dearly say that
9=P 9(H(P1) + 9rH(Pr) (H(P)
9:00
Hence, average enmopy spir is no greater
Hence, average enmopy spect is no greature than original enmopy at mode.

Answer 6b

Ans.66)	$H(CIT) = \sum_{c} P(C_{c}, T_{c}) \log P(C_{c}, T_{c}) $
	=> EE P(Ce,TE) log P(Cc TE)
	Marginalise all over classes Σ PCCc) = 1
	⇒ E P(T _E) P(C _C T _E) log(P(C _C T _E))
	=> E P(Tt) H(Co Tt)
	=> E P(T=t) H (CIT=t)
	This is RHS of given equation to prone.
t	ve Started with LHS and prone that to be equal to RHS. Given Equation:
	Ginen Equation:
	H(CIT) = ZP(Tzt) H(CITzt)

Answer 6c

MIISWEI OC	
Ans GC	To prone: H(T,C) = H(CIT) + H(T)
W. Q. Q.C.)	10 proves
	Paus rule: P(x,y) = P(4) + P(x)
	Bayes rule: P(x,y) = P(y) + P(x)
	H(T() = H((T) + H(T))
	so let us write $H(T,C) = H(CTT) + H(T)$ in terms of probabilities and try to the use bayes rule on it.
	in terms of probabilities and
	use bayer rule on it.
	How y purp & drive & sound forming for
	$H(x,y) = \sum_{x=y}^{p} \sum_{x=y}^{p} P(x,y) \log \frac{1}{P(x,y)} (tormula tor py)$
	TI) H top our
	2 Z z P(4)x) · P(x) · log 1
	2 Z z P(y x) · P(x) · log // P(y x) · P(x)
	ting proved
	$= 7 \sum_{x \in P(y x)} P(x) \left[\log \left(\frac{1}{P(y x)} \right) + \log \left(\frac{1}{P(x)} \right) \right]$
	x y P(y x) (P(x))
	2) 5 5 10/11/20 , 0/4 / 100 / 1 / 1 5 5 0/11/20 / 20/1/20
	=> ZZ (P(y x)·P(x)·log (1) + ZZ P(y x)·P(x)·log
	(((((((((((((((((((
	=> EP(x). EP(Y x). log (P(y x)) + Z P(x). log (P(x)). EP(Y x)
	2 ((P(y)x) 2 ((P(x)) g
	=> \ \(\(\) \ \(\) \ \(\) \ \(\) \ \(\) \ \(\) \ \(\) \ \(\) \ \(\) \ \(\) \ \(\) \ \(\)
	2 (Pa)
	$\alpha = \sum_{y} p(y x) = 1.$

THE RESIDENCE OF THE PARTY OF T	16 prove: H(T,C) =
	The state of the s
= P(418) + RED)P(2C)	Bayes rule 1 P(x,y)
so equation ()	cauld be written as
(T) H+ (TI) H= (),T)	so let us mith H
H(X,Y) => H(Y X) +	Ha(x) of 10 2mmst in
+1 N	use payer rule o
Now replace, 4	with C and X with
	H(x,4) = \$ = P(x,4)
(hours (hood)	
we get H(CIT
H(T,C)	= 19 H (dta) + H (t
(x)9 (yi)9 V	0
Hence, proved.	