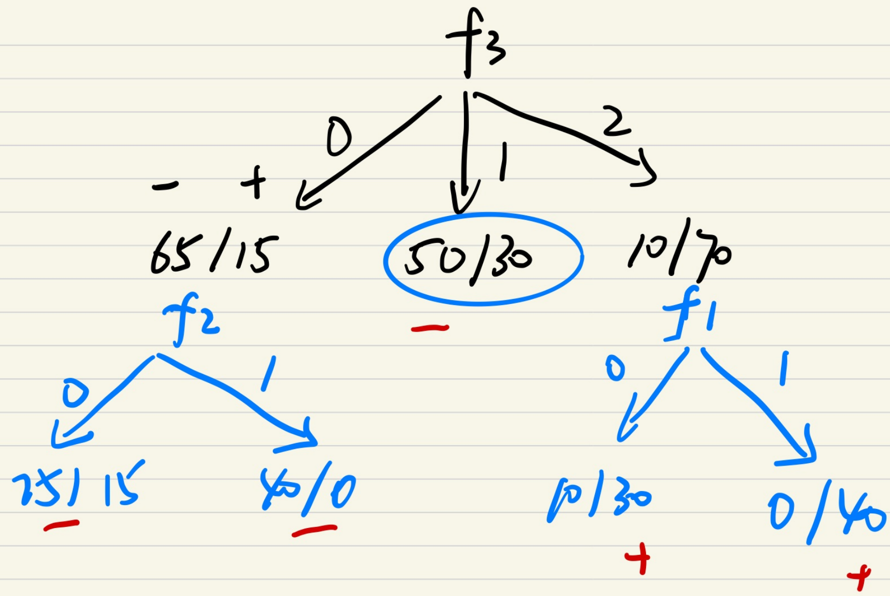
Homework 05

Dominik Sellenthin 2836308
Min Wang 3440557
Sisi Miao 3377882

1 Inductive Construction

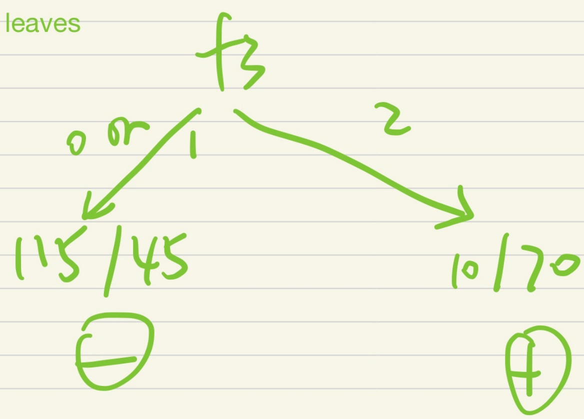
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	f1	f2	f3	f4	-	+			1st	Split by f1					
2	0	0	0	0	10	0			Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
3	0	0	0	1	5	5			0	0.5	70	50	0.583333333	0.416666667	0.979869
4	0	0	1	0	10	0			1	0.5	55	65	0.458333333	0.541666667	0.994985
5	0	0	1	1	10	0			H(Y|X)=	0.987426792	IG(Y|X)=	0.011320506			
6	0	0	2	0	0	10									
7	0	0	2	1	0	10			1st	Split by f2					
8	0	1	0	0	10	0			Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
9	0	1	0	1	10	0			0	0.5	50	70	0.416666667	0.583333333	0.979869
10	0	1	1	0	5	5			1	0.5	75	45	0.625	0.375	0.954434
11	0	1	1	1	0	10			H(Y|X)=	0.96715138	IG(Y|X)=	0.031595918			
12	0	1	2	0	10	0									
13	0	1	2	1	0	10			1st	Split by f3					
14	1	0	0	0	0	10			Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
15	1	0	0	1	10	0			0	0.333333333	65	15	0.8125	0.1875	0.696212
16	1	0	1	0	5	5			1	0.333333333	50	30	0.625	0.375	0.954434
17	1	0	1	1	0	10			2	0.333333333	10	70	0.125	0.875	0.543564
18	1	0	2	0	0	10			H(Y|X)=	0.731403569	IG(Y|X)=	0.267343729			
19	1	0	2	1	0	10									
20	1	1	0	0	10	0			1st	Split by f4					
21	1	1	0	1	10	0			Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
22	1	1	1	0	10	0			0	0.5	70	50	0.583333333	0.416666667	0.979869
23	1	1	1	1	10	0			1	0.5	55	65	0.458333333	0.541666667	0.994985
24	1	1	2	0	0	10			H(Y|X)=	0.987426792	IG(Y|X)=	0.011320506			
25	1	1	2	1	0	10			After 1st split, Max IG(Y|X)=			0.267343729	So, we choose f3 as the root		
26	P(-)=	0.52083			Number(-)	Number(+)									
27	P(+)=	0.47917			125	115									
28	H(Y)=	0.99875													

2nd	When f3=0					
	Split by f1					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	35	5	0.875	0.125	0.543564
1	0.5	30	10	0.75	0.25	0.811278
H(Y|X)=	0.677421284	IG(Y|X)=	0.053982285			
	Split by f2					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	25	15	0.625	0.375	0.954434
1	0.5	40	0	1	0	0
H(Y|X)=	0.477217001	IG(Y|X)=	0.254186567			
	Split by f4					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	30	10	0.75	0.25	0.811278
1	0.5	35	5	0.875	0.125	0.543564
H(Y|X)=	0.677421284	IG(Y|X)=	0.053982285			
After 2nd split, Max IG(Y|X)=			0.254186567	So, we choose f2 as the split feature		
2nd	When f3=1					
	Split by f1					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	25	15	0.625	0.375	0.954434
1	0.5	25	15	0.625	0.375	0.954434
H(Y|X)=	0.954434003	IG(Y|X)=	-0.223030434			
	Split by f2					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	25	15	0.625	0.375	0.954434
1	0.5	25	15	0.625	0.375	0.954434
H(Y|X)=	0.954434003	IG(Y|X)=	-0.223030434			
	Split by f4					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	30	10	0.75	0.25	0.811278
1	0.5	20	20	0.5	0.5	1
H(Y|X)=	0.905639062	IG(Y|X)=	-0.174235493			
After 2nd split, Max IG(Y|X)=			-0.174235493	Since the max IG<0, we make f3=1 as leaf node.		
2nd	When f3=2					
	Split by f1					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	10	30	0.25	0.75	0.811278
1	0.5	0	40	0	1	0
H(Y|X)=	0.405639062	IG(Y|X)=	0.325764507			
	Split by f2					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	0	40	0	1	0
1	0.5	10	30	0.25	0.75	0.811278
H(Y|X)=	0.405639062	IG(Y|X)=	0.325764507			
	Split by f4					
Vj	P(X=Vj)	Num(Y=-|X=Vj)	Num(Y=+|X=Vj)	P(Y=-|X=Vj)	P(Y=+|X=Vj)	H(Y|X)
0	0.5	10	30	0.25	0.75	0.811278
1	0.5	0	40	0	1	0
H(Y|X)=	0.405639062	IG(Y|X)=	0.325764507			
After 2nd split, Max IG(Y|X)=			0.325764507	Since the IG are the same for the rest features, we can just choose one of them.		



$$\frac{15+0+30+10+0}{125+125} = 22.9\%.$$

After merge the leaves



For the detail of calculation, see the excel document.

2 Minimal Error Pruning

$$\frac{24+9+2}{2948+2738} = 0.616\%$$

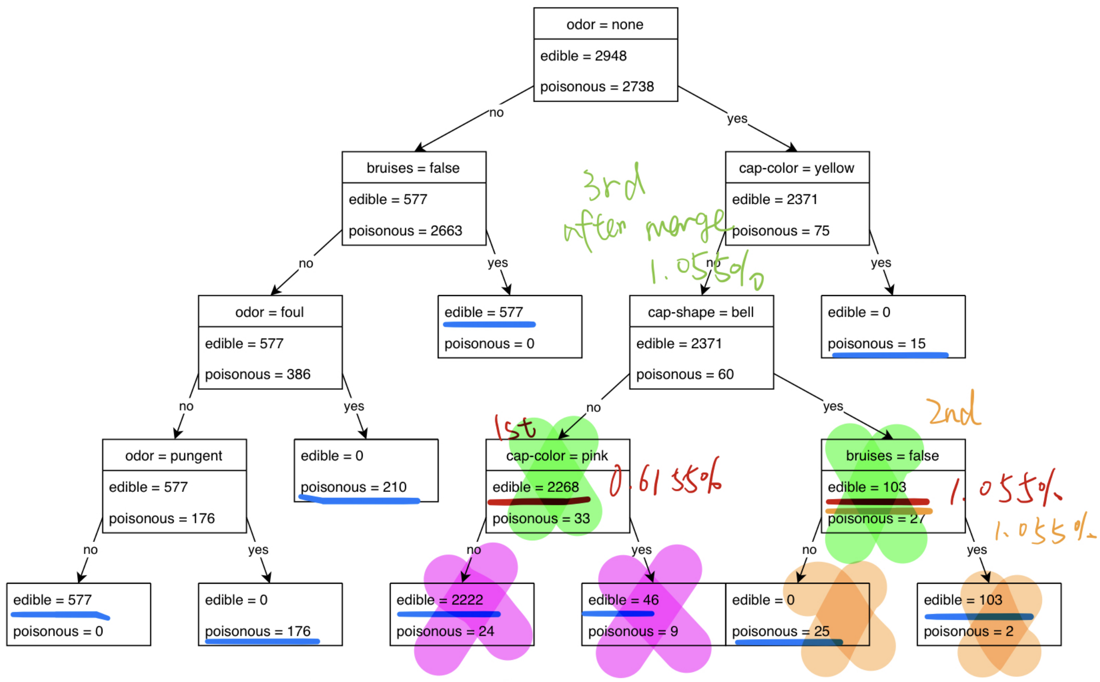


Figure 1: Decision tree of depth 4 classifies mushrooms as edible or poisonous. Inner nodes have a decision rule at the top, the left child is chosen if the rule does not apply and the right child is chosen if the rule applies. For each leaf and inner node, the class frequencies for the training data are summarized.

We use different color to donate the tree of original (black), after 1st pruning (red), 2nd pruning (orange) and after merging (green).

3 Regression with Decision Tree and KNN

1. Each leaf of Regression tree is numeric value ($y_{prediction}$ corresponds to the average y_{true} from different cluster of observations), in contrast, classification tree have true/false or discrete category as leaves.

The prediction computed as follows:

Decide root: Split the data into two groups by find the threshold that give us the smallest sum of squared residuals.

Steps :

(a) Rank the feature data in ascending order.

(b) Calculate the average of adjacent feature data. The average will be a candidate threshold to split the data.

(c) Calculate the squared residuals for each threshold. Choose the threshold with lowest squared residuals as split point of this numeric feature.

Decide nodes and leaves:

Repeat the above steps to split the data further. Note: To avoid overfitting, we only split observations when they are more than the setting minimum (i.e., if not, we make it as a leaf.) Typically the minimum number of observations to allow for a split is 20. If the data doesn't have many observations, we can set a smaller minimum.

2. In contrast to classification problem, in which we picks the greatest vote label as predicted class among k most close neighbors, the average of the y_true of k neighbors is taken to be the final prediction in regression KNN.

For further references see Regression Trees, Clearly Explained and A Practical Introduction to K-Nearest Neighbors Algorithm for Regression.