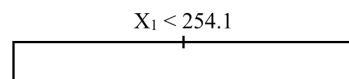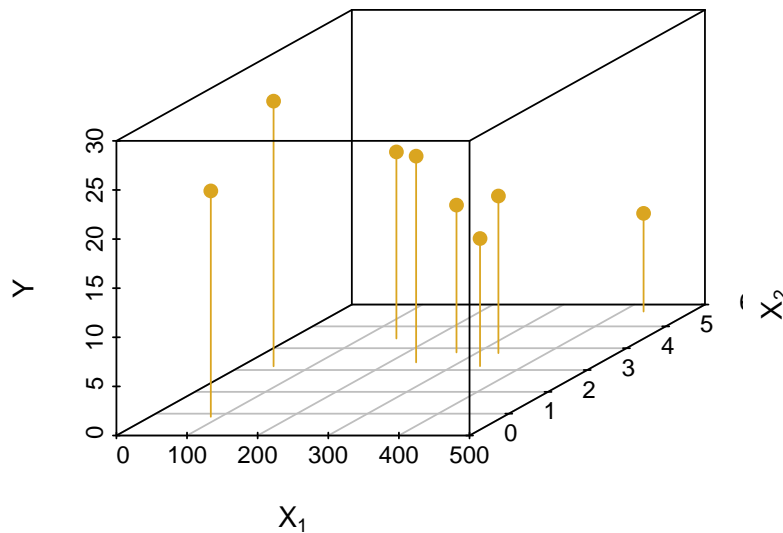## Boosted Trees

One Weak Learner $(b = 1)$: We'd like to build a boosted tree that will predict a value for Y given X_1 and X_2 using the 8 training observations found in the table below.

1. The structure of the first weak learner is shown in the tree diagram below. Please fill in the prediction at each of the leaves as well as in the table in the column labeled $\hat{f}^1(x)$.

$X_1 < 254.1$

	X_1	X_2	Y	$\hat{f}^1(x)$	r
1	270.00	1.25	15.00		
2	236.03	1.75	21.00		
3	60.93	2.94	27.00		
4	255.95	4.38	15.00		
5	46.59	1.87	30.00		
6	108.66	1.48	26.00		
7	407.33	6.58	15.00		
8	478.19	4.32	19.00		

2. Carefully draw the surface corresponding to $\hat{f}^1(x)$ on the 3D scatterplot below.


3. Now that you have the prediction from the weak learner, please compute the residuals for each observation to find out where this learner did well and where it did poorly. Fill them in the column labeled r.
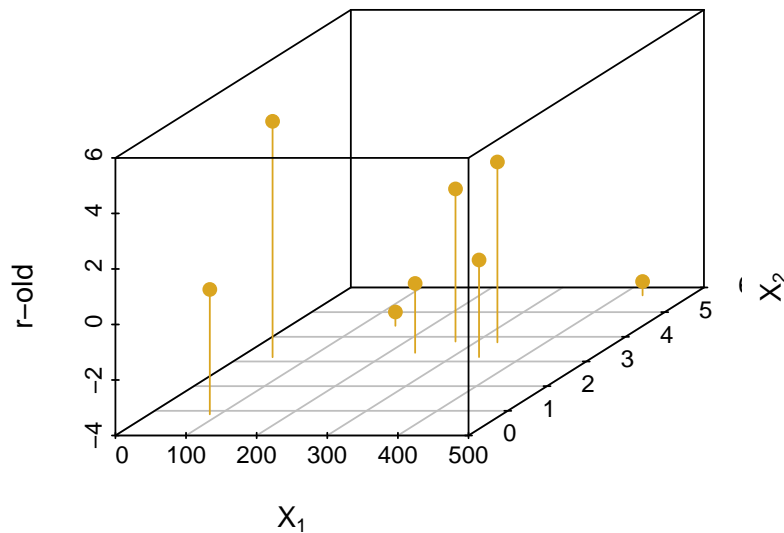
Boosted Trees

Second Weak Learner ($b = 2$): Now we strengthen our tree by incorporating a second weak learner that addresses the most glaring deficiencies of the first learner.

1. The structure of the second weak learner is shown in the tree diagram below. Please fill in the prediction at each of the leaves as well as in the table in the column labeled $\hat{f}^2(x)$.

$X_2 < 4.1$
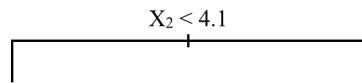
	X_1	X_2	r_{old}	$\hat{f}^2(x)$	r_{new}
1	270.00	1.25	-1.00		
2	236.03	1.75	-5.00		
3	60.93	2.94	1.00		
4	255.95	4.38	-1.00		
5	46.59	1.87	4.00		
6	108.66	1.48	0.00		
7	407.33	6.58	-1.00		
8	478.19	4.32	3.00		

2. Carefully draw the surface corresponding to $\hat{f}^2(x)$ on the 3D scatterplot below.


3. Now that you have the prediction from the second weak learner, please compute the new residuals for each observation to find out where this learner did well and where it did poorly. Fill them in the column labeled r_{new}.
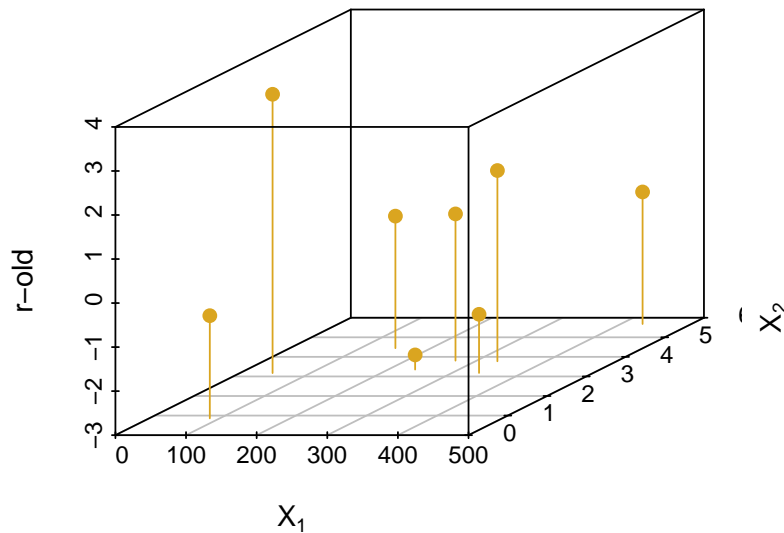
Boosted Trees

Third Weak Learner ($b = 3$): We continue to strengthen our tree by incorporating a third weak learner that addresses the most glaring deficiencies of the first two learners.

1. The structure of the third weak learner is shown in the tree diagram below. Please fill in the prediction at each of the leaves as well as in the table in the column labeled $\hat{f}^3(x)$.

$X_1 < 117.8$
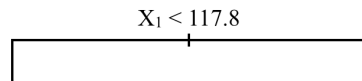
	X_1	X_2	r_{old}	$\hat{f}^3(x)$
1	270.00	1.25	-0.80	
2	236.03	1.75	-4.80	
3	60.93	2.94	1.20	
4	255.95	4.38	-1.33	
5	46.59	1.87	4.20	
6	108.66	1.48	0.20	
7	407.33	6.58	-1.33	
8	478.19	4.32	2.67	

2. Carefully draw the surface corresponding to $\hat{f}^3(x)$ on the 3D scatterplot below.

Boosted Trees

Full Boosted Model: Let's say we're sufficently satisfied with having three weak learners. The final step is to combine them into the final model.

1. Fill in the final boosted estimates, $\hat{f}^{boost}(x)$, in the table below.

	X_1	X_2	Y	$\hat{f}^1(x)$	$\hat{f}^2(x)$	$\hat{f}^3(x)$	$\hat{f}^{boost}(x)$
1	270.00	1.25	15.00	16.00	-0.20	-1.12	
2	236.03	1.75	21.00	26.00	-0.20	-1.12	
3	60.93	2.94	27.00	26.00	-0.20	1.87	
4	255.95	4.38	15.00	16.00	0.33	-1.12	
5	46.59	1.87	30.00	26.00	-0.20	1.87	
6	108.66	1.48	26.00	26.00	-0.20	1.87	
7	407.33	6.58	15.00	16.00	0.33	-1.12	
8	478.19	4.32	19.00	16.00	0.33	-1.12	

2. Carefully draw the surface corresponding to $\hat{f}^{boost}(x)$ on the 3D scatterplot below (yes, this is difficult!). Note that we're back to the same coordinates as in the scatterplot for the first weak learner.
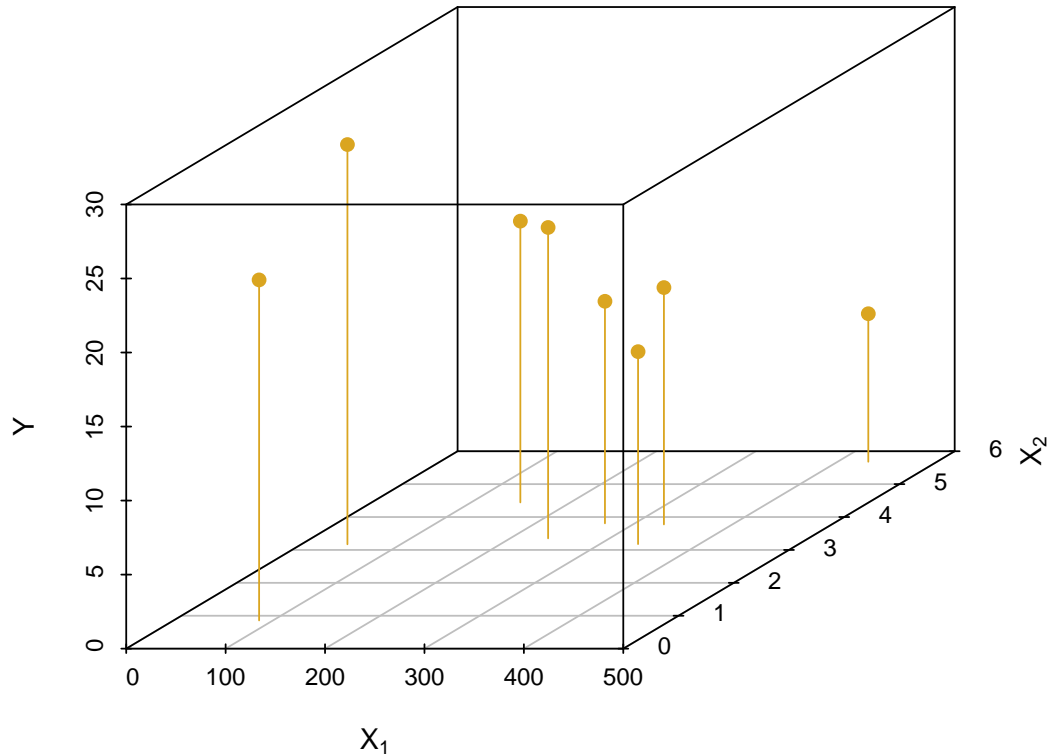
3. What value did we use for d, the depth of each weak learner?

4. What value did we use for B, the number of weak learners?

5. What value did we use for λ, the shrinkage parameter?

6. Please describe the trend in the amount of explanatory power of each successive learner, $\hat{f}^1(x)$ to $\hat{f}^3(x)$.

7. What would be the consequences if we set $\lambda = 0.01$?