Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

ML Interpretation

TOTAL POINTS 6

1. You train the random forest pictured below and it gets a c-index of 0.90. After shuffling the values for x, your dataset is the following. What is the variable importance for x? **1 / 1 point**

ID	x	y	death
1	2	3	1
2	4	5	0
3	1	2	1
4	5	2	0

- 0.1
- 0.5
- (selected) 0.65

Correct
Explanation: We need to calculate the new C-index. The prediction for 1 is low risk, the prediction for 2 is low risk, the prediction for 3 low risk, and the prediction for 4 is high risk. The permissible pairs are (1, 2), (1, 4), (3, 2), (3, 4). All of these are risk ties except for (3, 4) and (1, 4), which are not concordant. Therefore the c-index is 0.5(2) / 4 = 0.25. Therefore the difference between the original C-index and the new one is 0.9 - 0.25 = 0.65, so the answer is D.

2. Say you have trained a decision tree which never splits on a variable X. What will be the variable importance for X using the permutation method? **1 / 1 point**

- 0.5

what metric we use the variable importance will be 0, since there will be no change in the model output. Therefore the answer is C.

3. We have the following table the output of a model f on an example using subsets of the variable. What is the Shapley value for s_BP? **1 / 1 point**

Feature Set	Output
{}	0.5
{s_BP}	0.7
{d_BP}	0.6
{s_BP, d_BP}	0.65

- 0.0
- 0.2

{d_BP, **s_BP**} - {d_BP} = (0.65) - (0.6) = **0.05**

{**s_BP**} - {} = (0.7) - (0.5) = **0.2**

Once we have obtained all of our values, we sum them up altogether, then divide by the number of features we have. In this case, we have 2 total features, so we divide by 2.

Calculate the importance of **s_BP**:

((0.05) + (0.2)) / 2

(0.25) / 2

The shapley value for s_BP is: **0.125**

4. We have the following table the output of a model f on an example using subsets of the variable. What is the sum of the Shapley value for s_BP and d_BP? **1 / 1 point**

Feature Set	Output
{}	0.5
{s_BP}	0.7
{d_BP}	0.6
{s_BP, d_BP}	0.65

We already know the Shapley value of **s_BP** from Question 3 (0.125). Thus, all we need to calculate is the Shapley value from **d_BP**.

We compute the shapley value for d_BP in the following way:

{s_BP, **d_BP**} - {s_BP} = (0.65) - (0.7) = **-0.05**

{**d_BP**} - {} = (0.6) - (0.5) = **0.1**

Once we have obtained all of our values, we sum them up altogether, then divide by the number of features we have. In this case, we have 2 total features, so we divide by 2.

Calculate the importance of **d_BP**:

(0.1 + (-0.05)) / 2

(0.05) / 2

The Shapley value for d_BP is: **0.025**

Since we want to calculate the sum of the Shapley value for s_BP and d_BP, and we already know the value of s_BP from the previous exercise we can sum:

{**s_BP**} + {**d_BP**} =

(0.125) + (0.025) =

0.15

- (selected) No

Correct
Explanation: The answer is no. We see that when only adding d_BP, the output goes up, so the coefficient for it must be positive. We also see that when only adding s_BP the output increases, so the coefficient must be positive. However, when we add d_BP to the output with s_BP, the output goes down, a contradiction, since we already know the coefficient for d_BP is positive. This suggests that there must be at least an interaction between s_BP and d_BP.

6. Now assume we add Age as a variable. What is the new Shapley value for s_BP? **1 / 1 point**

Feature Set	Output
{}	0.5
{s_BP}	0.7
{d_BP}	0.6
{Age}	0.7
{s_BP, d_BP}	0.65
{s_BP, Age}	0.7
{d_BP, Age}	0.8

- (selected) 0.09
- 0.125