

7/9 Questions Answered

Saved on Dec 4 at 5:20 PM

HW5 Conceptual

Q1 Interpretability of Random Forest (RF) Models

30 Points

We know that we can write a decision tree as a set of rules, which makes trees one of the most interpretable models in Machine Learning. We also know that a forest of trees leads to more powerful predictive models, alas at the expense of interpretability.

Your mission is to **propose an algorithm** that can extract "**feature importance**" from a RF model.

Discuss the pros and cons of your algorithm.

It is fine to research the topic, but try to propose your own solution.
Please provide all your sources.

Compelling answers with clear steps will receive full credit.

Hint: Feel free to use functions (entropy, gini, etc.), data, manipulate the feature values, whatever you think is useful to identify those important features.

Please show your work.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

 Please select file(s)

Select file(s)

选择文件 未选择任何文件

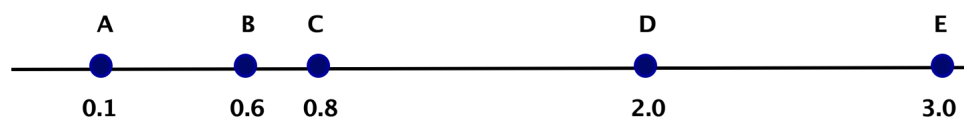
Enter your answer here

Save Answer

Q2 K-means clustering

30 Points

Consider the one-dimensional dataset represented below.



We would like to run **K-Means** clustering for $k = 2$.

Suppose you initialize the two centers to $\mu_1 = A$ and $\mu_2 = B$

Write down the cluster assignments at each iteration along with the centers.

Show all your steps.

Please show your work.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

 Please select file(s)

Select file(s)

0 : $\mu_1 = 0.1$ $\mu_2 = 0.6$. --> $\mu_1' = 0.1$ $\mu_2' = (0.2 + 1.4 + 2.4)/4 = 1$
 1 : $\mu_1 = 0.1$ $\mu_2 = 1$ --> $\mu_1' = 0.1$ $\mu_2' = (0.4 + 0.2 + 1 + 2)/4 = 0.9$
 2 : $\mu_1 = 0.1$ $\mu_2 = 0.9$ --> $\mu_1' = 0.1$ $\mu_2' = (0.3 + 0.1 + 1.1 + 2.1)/4 = 0.9$

Save Answer

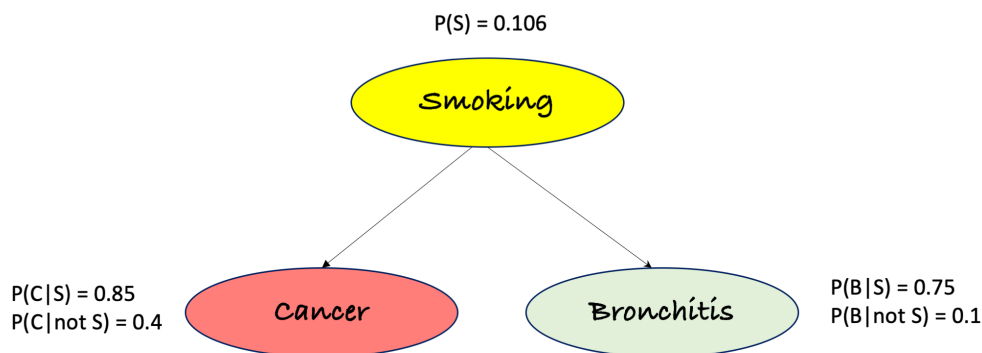
Last saved on **Dec 03 at 5:37 PM**

 选择文件 未选择任何文件

Q3 Bayes Nets: Smoking is bad!

40 Points

A doctor is modeling his beliefs about the relationship between smoking (S), lung cancer (C), and bronchitis (B). He drew the following Bayes network.



We will use binary random variables. To make notations easy, we will use random variables or their negation to represent when the variable is True or False respectively.

Let S denotes smoking is true and \bar{S} denotes smoking is False. Likewise, we will denote by C and \bar{C} for Cancer and no cancer resp., B and \bar{B} for bronchitis and no bronchitis resp.

Please show your work.

Q3.1

8 Points

Write the full joint distribution over the variables S , C , and B .

Hint: calculate $P(S, C, B), P(S, C, \bar{B}), \dots, P(\bar{S}, \bar{C}, \bar{B})$.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

CURRENTLY UPLOADED FILES

▼ png.png

Download Remove

选择文件 未选择任何文件

$$P(\bar{S}, C, B) = 0.106 * 0.85 * 0.75 = 0.067575$$

$$P(S, C, \bar{B}) = 0.106 * 0.25 * 0.85 = 0.022525$$

$$P(S, \bar{C}, B) = 0.106 * 0.15 * 0.75 = 0.0119$$

$$P(\bar{S}, C, B) = 0.894 * 0.4 * 0.1 = 0.03576$$

$$P(S, \bar{C}, \bar{B}) = 0.106 * 0.15 * 0.25 = 0.00397$$

$$P(\bar{S}, C, \bar{B}) = 0.894 * 0.4 * 0.9 = 0.32184$$

$$P(\bar{S}, \bar{C}, B) = 0.894 * 0.6 * 0.1 = 0.05364$$

$$P(\bar{S}, \bar{C}, \bar{B}) = 0.894 * 0.6 * 0.9 = 0.48276$$

 Please select file(s)

Select file(s)

Enter your answer here

Save Answer

Last saved on **Dec 03 at 9:13 PM**

Q3.2

8 Points

Calculate $P(C)$ using the full joint distribution.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

 Please select file(s)

Select file(s)

$$P(C) = 0.067575 + 0.022525 + 0.03576 + 0.32184 = 0.4477$$

选择文件

未选择任何文件

Last saved on **Dec 03 at 9:15 PM**

Q3.3

6 Points

Calculate $P(\bar{C}, B|S)$ (Hint: use conditional independence).

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

CURRENTLY UPLOADED FILES

▼ png (1).png

Download Remove

$$P(\bar{C}, B|S) = P(\bar{C}|S) * P(B|S) * P(S) = 0.15 * 0.75 * 0.106 = 0.0119$$

Please select file(s)

Select file(s)

Enter your answer here

Save Answer

Last saved on Dec 03 at 9:20 PM

Q3.4

6 Points

Calculate $P(S|\bar{C}, B)$ (Hint: use Bayes rule).

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

CURRENTLY UPLOADED FILES

▼ png-1.png

Download Remove

$$P(S|\bar{C}, B) = 0.0119 * 0.106 / (0.0119 + 0.05364) = 0.019246$$

Please select file(s)

Select file(s)

选择文件 未选择任何文件

Enter your answer here

Save Answer

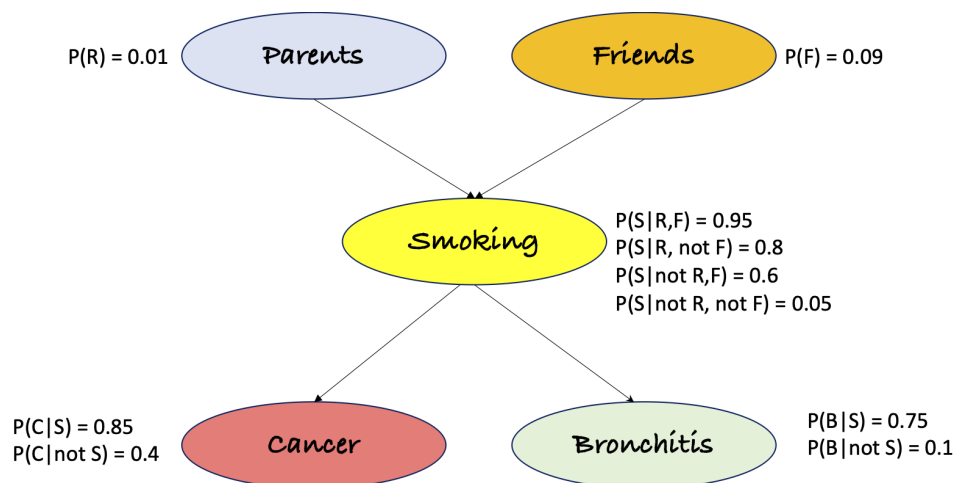
Last saved on **Dec 03 at 9:27 PM**

Q3.5

6 Points

The doctor figured out that there are two reasons for smoking. The influence of **friends (F)** during teen years (apparently, teens smoke to look as cool as their friends) and having **parents (R)** who smoke (apparently, kids inhale as much nicotine as their parents when these smoke, and nicotine is addictive). F and R are also binary variables.

He updated the BN to the following:



Using the chain rule, calculate $P(\overline{C}, B, S, R, \overline{F})$.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

Please select file(s)

Select file(s)

$P(\text{not } C, B, S, R, \text{not } F) = P(R) \cdot P(\text{not } F) \cdot P(S | R, \text{not } F) \cdot P(\text{not } C | S) \cdot P(B | S)$

选择文件 未选择任何文件

$$= 0.01 * 0.91 * 0.8 * 0.15 * 0.75$$

$$= 0.000819$$

Save Answer

Last saved on Dec 03 at 9:32 PM

Q3.6

6 Points

Same BN as the previous question. Show that Parents and Friends are independent variables.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

CURRENTLY UPLOADED FILES

▼ IMG_6A782006F04D-1.jpeg

Download ✕ Remove

$$P(P, F) = \sum_S P(P, F, S)$$

$$= \sum_S P(P) \cdot P(F) \cdot P(S|P, F)$$

$$\Rightarrow P(P, F) = P(P) \cdot P(F)$$

Thus, Parents and Friends are independent variables.

Please select file(s)

Select file(s)

Enter your answer here

Save Answer

Last saved on Dec 04 at 5:20 PM

Q4 Deadline

0 Points

Please make a note here if you received an extension from the teaching staff. If not, please leave this blank.

Please note this assignment only has a due date, there is no late date.

Enter your answer here

Save Answer

Save All Answers

Submit & View Submission >