In Class Activity – Naïve Bayes (ICA 11) - Solutions

Please enter your responses at https://tinyurl.com/AIF19-ICA11

The table below shows the current results of the CU football team. The next game is with Washington State. Your goal is to predict the outcome using a Naïve Bayes classifier.

CU Opponent	Home or away?	Time of day	Opponent ranked higher?	Win or loss
Colorado State	Home	Night	No	Win
Nebraska	Home	Afternoon	Yes	Win
Hollywood U	Home	Night	No	Loss
Air Force	Home	Morning	No	Loss
ASU	Away	Night	Yes	Win
Arizona	Away	Afternoon	Yes	Loss
Oregon	Away	Night	Yes	Loss
Washington St	*Away*	*Afternoon*	*No*	*?*

Please enter your responses to the following questions on the Google form.
- a) What is the probability of CU winning?
- b) What is the probability of CU losing?
- c) What is the predicted outcome of the game?
- d) How confident (on a 1 [not at all confident] to 6 [very confident]) are you in your prediction?
- e) How would the classifier be affected if CU had lost the ASU game?

Probabilities needed:
P(win) = 3/7
P(loss) = 4/7
P(Away | win) = 1/3
P(Away | loss) = 2/4
P(Afternoon| win) = 1/3
P(Afternoon | loss) = 1/4
P(Opp ranked higher = no | win) = 1/3
P(Opp ranked higher = no | loss) = 2/4

X = <Away, Afternoon, Opp. Ranked higher = No>
P(X| win) = P(Away | win) * P(Afternoon| win) * P(Opp ranked higher = no | win) = 1/3*1/3*1/3 = 0.037
P(X | loss) = P(Away | loss) * P(Afternoon | loss) * P(Opp ranked higher = no | loss) = 2/4*1/4*2/4 = 0.063

Solution to a) P(X | win) * P(win) = 0.037*3/7 = 0.016
Normalize to produce 0 − 1 probability:
$$\frac{P(X \mid win) * P(win)}{(P(X \mid win) * P(win) + P(X \mid loss) * P(loss)} = \frac{0.016}{0.016 + 0.036} = 0.308$$

Solution to b) P(X | loss) * P(loss) = 0.063*4/7 = 0.036
Normalize to produce 0 − 1 probability:
$$\frac{P(X \mid loss) * P(loss)}{(P(X \mid win) * P(win) + P(X \mid loss) * P(loss)} = \frac{0.036}{0.016 + 0.036} = 0.692$$

Solution to c): Prediction: CU Loses. Since P(X | loss) > P (X | win) we classify CU as likely to lose to Washington State in the upcoming game.

Solution to d): P(Away | win) would equal 0. This would set P(X | win) to equal 0. The classifier would then never predict CU as winning. This is known as a zero-sampling frequency problem in the dataset. The target situation has never occurred in the data, even though it is known to possibly occur, this means the dataset resulting probability doesn't represent the actual event probability due to insufficient sampling.