**Part 2 – General Questions**

Answer the following questions. Make sure to explain your responses.

1. A, B and C are events and P(X) denotes the probability of X. Also, A implies B and B implies C.
   1. P(A) ≤ P(C).
   2. P(A) ≥ P(C).
   3. P(B ⋂ C) ≥ P(B).
   4. P(B ⋂ C) ≤ P(B).
   5. More information is needed to decide whether c) and d) hold or not.

**D is true**

**If B implies C, that tells us that if B then C. We know that events B and C are not mutually exclusive. If they were disjoint we know that P(B ⋂ C) = 0.**

**If B were contained entirely within C then P(B ⋂ C) = P(B)**

**If C were contained entirely within B then P(B ⋂ C) = P(C) where P( C) <= P(B)**

**Therefore P(B ⋂ C) <= P(B)**

1. What can you say about this formula ?

**The result of this series approaches infinity and does not have a finite limit.**

**As an example –**

= 1 + + + + + …

**We can see that the terms are decreasing as n increases.**

**For n=2**

**S2=** 1+

**S4 =** 1+ + +

**S8 =** 1+ + + + + + +

**If you examine S4, it can be rewritten as**

**S4 >**1+ +

**because**  + **is greater than**

**Likewise, S8 can also be rewritten:**

**S8 >** 1+ +

**because the final four terms of S8 are also greater than**

**Put another way,**

**S2=** 1+  **,**

**S4 >**1+ ,

**S8 >**1+

**S16 >**1+

**And so on, such that we can generalize Sn for**

**S2^n >** 1+

**As n increases, the term being added to 1 keeps increasing and therefore the result of the series approaches infinity and is a divergent series.**

1. A magician has two cards: one is white on both faces and the other is black on one side and white on the other. The magician picks one card at random (with probability 1/2 for each) and throws it on the floor in a way that does not favor a card side. The card on the floor is white on its visible side. **What is the probability that the other side is white as well?**

**2/3 – there are 4 faces, 3 are white. If I see one white face, I know that there are only 3 faces left, of which 2 could be white and 1 black. Therefore, 2/3.**

**Using conditional probability**

P (Wback| Wfront) = P (Wfront ⋂ Wback) / P (Wfront)

P (Wback| Wfront) = **/**

P (Wback| Wfront) =

1. For each of these statements about binary classification and the various measures related to confusion matrices, indicate if you agree or not, and explain why:
   1. If two models have different accuracy values, the one with a higher accuracy is better.
   2. Type II errors are less problematic than type I errors.
2. **Not necessarily – it depends on the problem we are working on. Accuracy is the measure of correctly predicted observations to the total observations – which is great if that’s all we care about. Sometimes we need to get more granular. In working with healthcare data, sometimes you care more about minimizing the number of false positives (for example, predicting that someone has cancer when they do not). In these cases you could perhaps look to precision as a better measure.**

**In working with a online casino, we wanted to predict future VIPs. Here we did not really care if we had false positives (predicting someone to be a VIP when they weren’t). Instead what we DIDN’T want was to let a potential whale slip through our fingers. We therefore could tolerate false positives but could not tolerate false negatives (predicting someone not to be a VIP who ended up being one). In this case Sensitivity was more important.**

1. **A type I error is when one rejects the null hypothesis when it is actually true while a type II error is failure to reject the null hypothesis when it is actually false.**

**To illustrate**

|  |  |  |
| --- | --- | --- |
|  | **Null hypothesis true** | **Null hypothesis false** |
| **Reject null hypothesis** | **Type I error** | **Great!** |
| **Fail to reject null hypothesis** | **Great!** | **Type II error** |

**Traditionally type I errors are more problematic.**

**Type I errors mean you are attributing statistical significance to findings that have no basis in reality. This could cause you to develop strategies and commit resources to something that will not pan out in reality.   
  
Type II errors would mean that you stick to the status quo. This could be less damaging from a sense of committing resources and organizationally, there’s less effort involved in turning the direction of a ship that’s already in motion.**

**However, at the end of the day you have to decide which error is more or less significant than the other – which one is more costly to you?**

**We could apply this thinking to a potential covid-19 vaccine.**

|  |  |  |
| --- | --- | --- |
|  | **Vaccine does NOT eradicate COVID19** | **Vaccine does eradicate COVID19** |
| **Reject null hypothesis** | **Type I error – we would have a potential vaccine that is a dud** | **Great!** |
| **Fail to reject null hypothesis** | **Great!** | **Type II error – we throw away the potential life saving medicine** |

**In this scenario, it is not as straightforward as saying that Type II errors are less problematic. It really depends on the context. If you ask me, in a situation where lives are at stake, it is likely worse to throw away the vaccine than to push through a vaccine that does not work, given that there are many vaccines in development.**

1. As part of a game’s development process, we invite 32 subjects randomly selected from a list of volunteers to play the game in its current state and ask them to score the game on a scale of 0 to 10 on 20 different criteria (movement, camera, shooting, story, difficulty, etc.). We repeat the same process one month later with the same subjects and the new current state of the game. Among the t-tests performed independently on each of the 20 score differences, one of them (the movement score) detected a positive significant difference in score at the 0.05 level. We report that the new version of the game displays significant improvements regarding character movement appreciation. **What are the potential flaws affecting the internal and external validity of this study?**

**Internal validity is the extent to which your results are trustworthy and not contaminated. How confident are we that what we are observing is true.**

**There are a couple of things to note. How can we be sure that we are observing an improvement when using the same group? What if all we are actually observing is a group of subjects that have gained familiarity with the controls and that is the only thing that causes a positive difference in the movement score? It is entirely possible that the more familiar someone is with a game, the more satisfied they are with how it plays.**

**External Validity is the extent to which your results can be generalized. Because the study used the same subjects there is an obvious flaw in generalizing the study to other subjects. Put simply, just because the same group reports an improvement does not mean that another will.**

**Besides that it is possible that the sample is not representative of the eventual target population. It’s possible your list of volunteers are drawn from different demographics than your target demo.**

1. A model for detecting cheater accounts in an online game is tested on 1000 thoroughly verified accounts containing 20 cheaters and 980 non-cheaters. The model flags 25 cheater accounts, 5 of which are actual cheaters. **Compute the following values associated with the confusion matrix**: Accuracy, Precision, Recall.

|  |  |  |
| --- | --- | --- |
|  | **Actual Cheaters** | **Non-Cheaters** |
| **Identified as Cheaters** | **5** | **20** |
| **Identified Non-Cheaters** | **15** | **960** |

**Accuracy = 5 + 960 / 1000 = 96.5%**

**Precision = 5/ 5+ 20 = 20%**

**Recall = 5 / 5+ 15 = 25%**

**This is an example of another situation where accuracy would fail us as a metric by which to judge performance, and precision or recall would be better. This is often in the case in situations when the data is unbalanced between the different classes.**

1. Write a pseudo-code chunk that finds the maximum of two numbers. You cannot use if-else or any other comparison operator.

**def findmax2num(x,y):**

**nums =[x,y]**

**nums.sort()**

**return nums[1]**

**Approach: define a list that takes the inputs (i.e. two numbers), sort the list and then output the second element in the list, which should be the max of the two numbers.**

1. Enumerate the tunable hyper-parameters of a feed-forward classification neural network with fully connected layers, a cross-entropy loss function with an elastic net regularization term using stochastic gradient descent with momentum and linearly decreasing learning rate.

**There are a whole host of tunable hyper-parameters including but not limited to:**

**# of hidden layers**

**# of neurons within those layers**

**Dropout**

**Type of activation**

**Technically the loss function is also a hyper-parameter. If the assumption is we are not married to cross-entropy then we could also try options such as mean squared error.**

**The same goes for gradient descent – if we are using momentum there are other optimization algorithms such as Adam and RMSprop. You can also play around with stochastic gradient descent vs mini-batch gradient descent and batch gradient descent.**

**The learning rate itself is a hyper-parameter, but so is the learning rate decay.**