*xxxxx*

*Khadija Khaldi, Romita Banerjee and Christoph F. Eick*

COSC 4368: Fundamentals of Artificial Intelligence Spring 2019

Problem Set3 (Individual Tasks) Version 4

Deadline: Su. April 28, **11p** (3% bonus); Wednesday, May 1, **11a** (the latest)

Available Points: 57

12. Logical Reasoning (4 points) Khadija

Show using Resolution (and **not** by using other methods!):

* + 1. xyz (P(x,y,z)  R(x,y) )
    2. rs (P(s,s,t)  Q(s,t) )
    3. ab (Q(a,a)  R(b,a) )
    4. xy (R(x,y)  R(y,y) )
    5. P(4,4,4)
    6. ~Q(4,5)

|-

(X) R(4,4)

First transform the FOPL formulas into clauses, and then the hunt for the empty clause can begin!

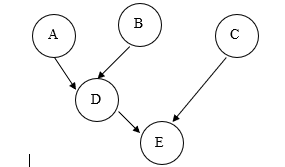
13. Bayes’ Theorem and Belief Networks (9 points) Romita



Fig. 1: Thomas Bayes 1740—279 years ago!

1. Assume we have 3 symptoms S1, S2, S3, a disease D and the following probabilities: P(D)=0.01 P(S1)=P(S2)=P(S3)=0.02; P(S1|D)=0.1; P(S2|D)=0.02; P(S3|D)=0.002. How would a naïve Bayesian system compute the following probability? **P(D|S1,S2,S3)**=…
2. Now assume the following additional knowledge has become available: P(S1,S2)=0.0002; P(S3|S1,S2)=0.08; P(S1,S2,S3|D)=0.000032; how would you use this information to obtain a “better” estimation of P(D|S1,S2,S3)?
3. How can the discrepancy with respect to the obtained probabilities between cases a) and b) be explained? Why are the numbers you obtain different? What does this discrepancy tell you about naïve Bayesian systems in general?

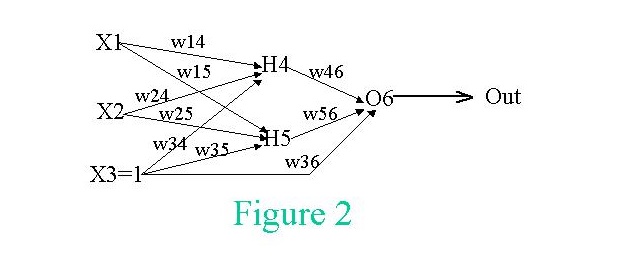
d) Assume that the following belief network is given that consists of nodes A, B, C, D, and E that can take values of true and false.



* Is A**|**D,B d-separable from C**|**D,B? Give reasons for your answer!
  + Yes, A and C are conditionally independent.
* Is A,E| d-separable from C|? Give reasons for your answer! :=”no evidence”

14. Weight Learning Computations in Neural Networks (7 points) Khadija

Assume we a 3-layer NN that is depicted in Fig. 2 is given that has 3 inputs X1, X2, and X3.



A training set D is given that consists of the following examples: (1,1,1,1) and (0, 0, 1, 0), indicating values for X1, X2, X3 and the “desired” activation for O6. Assume that the initial/current weights are w14 = 0.2, w15 = 0.2, w24 = 0.2, w25 =l 0.2, w34 = 0.2, w35 = 0.2, w36 = 0.5, w46 = 0.5, and w56 = 0.5. Use =0.5 as your learning rate. Show what new weights will be obtained after processing the two training instances of data set D. Use g(x) = 1/(1+e\*\*(-x)) as the activation function; that is: g'(x)=(e\*\*(-x))/(1+e\*\*(-x))\*\*2). List all important computations that lead to the reported weight updates.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S#** | **X1** | **X2** | **X3** | **O6** | **Des-O6** | **Error** | **w14** | **w15** | **w24** | **w25** | **w34** | **w35** | **w36** | **w46** | **w56** |
| **0** |  |  |  |  |  |  | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.5** | **0.5** | **0.5** |
| **1** | **1** | **1** | **1** |  | **1** |  |  |  |  |  |  |  |  |  |  |
| **2** | **0** | **0** | **1** |  | **0** |  |  |  |  |  |  |  |  |  |  |

15. Using a Belief Network Tool (20 points) Khadija



Fig. 3: Multiple Astronomers Looking at the Sky

Assume we have 3 astronomers in different parts of the world who make measurements M1, M2, and M3 of the number[[1]](#footnote-2) of stars N in some region of the sky. Normally, there is a probability of 0.05 that the astronomer counts a single star twice (overcounts by one star; you can assume that the three astronomers never undercount; moreover, if there is no star visible (N=0) the astronomer never overcounts). Moreover, there is a 10% probability (P(Fi=1)=0.1 for i=1,2,3) that a telescope is out of focus (represented using random variables F1, F2, and F3), in which the astronomer undercounts by 2 or more stars (e.g. if N is 4 and her telescope is out of focus, the astronomer will count 2, 1 or 0 stars; you can assume if information is missing that each case has the same probability). Design a belief network, and compute the probability of the other variables assuming the following pieces of evidence are given (feel free to use *Netica (*<http://www.norsys.com/download.html> ) or any another belief network tool to compute your answer[[2]](#footnote-3)!):

1. M1=3 M2=3 M3=1
2. M1=3 M2=3 M3=0
3. N=4, M2=1, M3=0
4. M1=0 M2=3 M3=2
5. N=3 F1=0 F2=0 F3=1
6. M1=4 M2=4 F3=1

Submit the complete Belief Network you created—including all its probability tables—, and the findings you obtained for the six cases listed above!

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **N** | **M1** | **F1** | **M2** | **F2** | **M3** | **F3** |
| 3 | **3** | InFocus | **3** | InFocus | **1** | OutOfFocus |
|  | **3** |  | **3** |  | **0** |  |
| **4** |  |  | **1** |  | **2** |  |
|  | **4** |  | **4** |  |  | **1** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

16. Ethical Problems of AI (17 points) Romita

Write an essay of at least 800 words, focusing on the Ethics and Governance of Artificial Intelligence Systems. Your essay should cover issues like balance between regulation and innovation, how AI is used to spread information, how to ensure that the AI systems follow our principles when making decisions and what responsibilities should they have among others.



Fig. 4: AI & Ethics

Be aware of the fact that plagiarism will not be tolerated in this course; however, this does not mean that you are not allowed to use material on the internet and taken from the scientific literature when writing your essay; you just need to cite the material you used and you will need to use quotations, if you use (parts of) sentences “unchanged” from other publications in your essay!

Some hopefully useful links for identifying a topic for your essay:

1. <http://web.stanford.edu/class/cs122/>
2. <https://www.vanderbilt.edu/strategicplan/undergraduate-residential-education/universitycourses-2018/ethics_of_artificial_intelligence.php>
3. <https://www.media.mit.edu/courses/the-ethics-and-governance-of-artificial-intelligence/>

Albert Einstein’s revelation that energy and matter are one in the same lead to experiments and engineering involving converting matter of one form into energy. This technology could be used for nearly by a nearly endless set of applications, but we know the first thing that it was used for – though nuclear reactors came eventually, the first application of this discovery was to end a significant amount of human life. Likewise the emerging technology of the day has the capability to either be used toward the advancement of human kind, or towards its destruction. Artificial Intelligence gives humans, mainly the creators of such intelligence, the power to influence the world in any way they please, whether for better or for worse. In the United States where development of these systems is significant, there is nearly no regulation regarding the control, use or creation of intelligence. In other nations leading in the advancement of AI, such as Russia and China, the government controls all aspects of the development of these systems and the common globally accepted status regarding these state’s government does not hold them in a high regard. Many great minds around the world fear the possible outcomes of our emerging technology, and others emphasize the possible positive outcomes of such technology. The time has come for mankind as a whole to determine the outcome of this technology – our global leaders must either step up to the table to take control of systems we barley understand or to allow them to increase in power until the eventual demise of humankind.

Since the ages of antiquity, man has long attempted to formalize reasoning and logic in such a way that the thoughts and mind of humans can be encoded in some mechanism. When Alan Turing published his Theory of Computation he took this idea to new levels, proposing that the simplest of system, that comprised of just 1’s and 0’s, would be able to simulate anything which was deemed computeable. Putting doomsday images of fire and destruction aside, the coming wave of automation threatens to leave a significant portion of the United State’s (and also, though most likely later, the world’s) population unemployed after their jobs are replaced by a machine that is able to do it better for less cost to the business. This is an outcome that is destined to take place regardless of the legislation regarding the development of and ethical methods used to create Artificially Intelligent systems. As CGP Grey’s video: *Humans Need Not Apply* suggests, the law of economics will take hold, and as these systems are developed and able to do the job better than any human, workers of flesh and blood will be replaced by those of electricity and steel. Though we must accept that this is the eventual outcome and that we have no control over whether or not it will happen, we must begin to consider what we will do when it does happen. As humans, we are defined by our work and our drive to be productive; a large component of our identify comes from our work – what we do is who we are. When particular human jobs begin to be phased out as robotic systems are created to replace them, we must have some plan for those humans who are phased out. Where will these people go? What sort of jobs could they be trained to do? Where will the money for this training come from? These are all questions that cannot be answered by an individual – they must be answered by the nations of the world. Our congress must step up to the plate and begin planning for this unavoidable outcome, before it is to late.

Companies like Facebook and Google make their revenue based on how long humans use their products, thus they attempt to maximize the time that humans stare at their illuminated screens, remaining constantly flooded with content and stimulation. Both of these companies have algorithms that attempt to generate, or select, content in such away that this human stimulation remains high. This is horrible for many reasons, most of which will not be covered in this essay. What will be commented on is the way in which this content is pushed to humans is completely based on the human themselves, it acts as a positive feedback loop: a person enjoys a thing so they find it on these sites, the sites then push more and more of this thing to the human who enjoys it more and more – hacking the human brain and taking advantage of innate human behavior common to all mammals. Take it while it is there, or in other words bingeing. This mechanism can be, and has been, exploited by individuals attempting to do significant harm to a significant amount of people. The 2016 United States Presidential Elections saw this occur in many regards, both from within its own country from its own political candidates as well as from overseas by agents attempting to influence our elections. These automatic content pushing systems allowed external individuals to display particular content to particular users; this lead targeted misinformation campaigns towards the individuals who were most inclined to believe it. Left wing liberals saw fake posts about how racist individuals were abusing minorities and how fascism was on a continual rise as rights were stripped from those who weren’t white men while right wing conservatives saw posts about how Europe was being overrun by ideologically driven Islamic fundamentalists imposing Sharia Law. As with most things in congress, legislation did not begin to be passed until this campaign was past its climax and the effects were already seen and felt around the nation.

Throughout the hearings that resulted from this ‘cyber warfare’, congress showed its utter ignorance regarding any and all things related to technology, especially the social media which now controls a large portion of our population. The driving question in the minds of many was: How can a who does not have even a basic understating of these technologies ever be hoped to regulate them? And from that I pose a further questions regarding many of the black box models we use to create AI systems: how can we ever hope to regulate systems that we do not understand?

For any hope of these systems to be created ethically the engineers creating them must constantly be keeping ethics within their mind. Companies cannot be result driven, at least not in the leading capitalistic notion. The possible impact to society for any groundbreaking technology or discovery must be high in the mind of the creators as a large majority of the population will not adequately understand it, or its effects. If an individual creates a technology sufficiently powerful, they may be able to influence the world in any way they please. For this reason to some extent legislation cannot help the issue, as a destructive technology could be crafted and exported into the world before legislative bodies even have a chance to consider it. Constant safety mechanisms must be in place at every point of the development process of AI systems – and beyond that the creator is the individual with the most power regarding how their technology will effect the world.

1. You can assume that N is limited to 4—but the astronomer do not know that: M1, M2 and M3 are therefore limited to values 0 through 5. [↑](#footnote-ref-2)
2. Including the answer ‘inconsistent’ in the case that the evidence is inconsistent, e.g, the evidence N=1 M1=3 is inconsistent—as it is ‘impossible’ because astronomer1 never overcounts by more than 1 star! [↑](#footnote-ref-3)