Assignment 5: HMMs, Bayesian Networks, and Decisions

Due: Wed Nov. 8, 11:59pm.

100 points.

For questions 1 and 4, please include a PDF with your answers.

Please also include a submission.py that can run your code for questions 2 and 3.

Question 1:

Our Mars rover has been out collecting samples, and it needs to return to the charging station as quickly as possible.

It knows that over rocky terrain it can go 2 km/h. Over sandy terrain it can go 3 km/h, and over smooth terrain it can go 5 km/h.

There are three routes it might choose from. Unfortunately, our terrain data for the three routes is incomplete, so we only have estimates.

Route 1 is 2 km long. There is a 20% chance it is sandy, 30% chance it is smooth, and a 50% chance it is rocky.

Route 2 is 1.8 km long. There is a 40% chance it is sandy, a 20% chance it is smooth, and a 40 % chance it is rocky.

Route 3 is 3.1 km long. There is a 50% chance it is sandy, a 40% chance it is smooth, and a 10% chance it is rocky.

(10 points) Which route should we pick? Show your work.

We have now found out some additional information.

Route 1 contains a crater. If the wall of the crater is intact, we can go through it. If the wall has been damaged, we will need to go around, which will add 45 minutes to our journey. There is a 30% chance that the wall is damaged.

Route 2 contains a bridge. If that bridge is damaged, we will need to repair it, which will add 1 hour to our time. There is a 60% chance that the bridge is out.

(10 points) Now which route should we pick? Show your work.

(10 points) Now suppose that we can use a satellite to find out whether the terrain in route 3 is smooth. Is this helpful? What is the value of this information? Expressed differently, how long are we willing to wait for this information from the satellite?

Having a satellite find out whether the terrain is route 3 is smooth or not would be helpful since that will help us determine whether we can take route 3 and go through it at the speed of 5 km/h.

However in this problem, charging and retaining the battery is the primary objective. So, depending on how much charge is left in the Rover's battery (assuming x time worth of battery for now). We could wait for up to X - 1.48 hours since that's most amount of time we could spend on route 1, which would then become current best path forward if route 3 is rocky.

(5 points) Now put this problem into ChatGPT. Is it able to solve it correctly? If not, where does it make mistakes?

If Route 3 is smooth, then we choose route 3 which takes 3.1*(1/5) = 0.62 hours. Sandy Route 3 will take 1.03 hours so we should pick route 1 Rocky Route 3, it will take 1.55 hours so we should also pick route 1

There is a 40% chance that it is smooth, and 60% otherwise so we have 0.62*(0.4) + 0.978*(0.6) = 0.83 hours

Before knowing this information, our EU was 0.92 hours, now it is 0.83 hours.

Thus, we are willing to wait up to 0.92 - 0.835 = 0.085 hours

I believe CHATGPT's math is really off along with it's logic. ChatGPT was still using expected value for the time taken to travel through route 3 but we can calculate the estimated time taken to travel through route 3 by considering the best and the worst case and thus estimate how much time is necessary for route 3.

ChatGPT says the following

```
the satellite information indicates that Route 3 is smooth (with a 40%
probability), the expected time for Route 3 would be reduced to
approximately 0.248 hours (the time for smooth terrain).
```

It incorrectly calculated the shortest possible time from 0.83 to 0.248.

(Problem 4 - grad students only)

AINow is a research institute that produces policy analysis addressing the concentration of power in the tech industry. They have recently published a landscape report assessing the state of the AI industry and making policy rcommendations.

Please read the executive summary and answer the following questions:

- What are the three dimensions along which Big Tech has an advantage in AI?
- AI has an advantage in AI because of the following reasons:
    1. Data - Big tech has access to large amounts of data, of all sorts of quality and types. This data is used to train AI models and make them more accurate.
    2. Compute - Big tech has access to large amounts of compute power, which is used to train and run AI models.
    3. Geopolitical power - Due to the large amount of data and compute power, Big tech has a lot of geopolitical power. This power is used to influence the direction of AI research and development through lobbying and other means. Additionally, due to this power, Big tech is able to embed itself into the defense and intelligence sectors of the government.
- Why does AI Now think it's important to focus on Big Tech?

- The author(s) of AI Now firmly believe that the direction and impact of AI can be mitigated and controlled if monitored actively. AI Now believes that the lifestyle of Big Tech, where they progress through moving fast and breaking things, cannot be allowed to affect the lives of people. AI Now is advocating to mandate monitoring and accountability of AI systems and with the increase in AI production, they believe that Big Tech needs to be monitored, now more than ever.

- Priority 1 discusses Algorithmic Accountability. What does this mean? Why is it important to shift responsibility for detecting harm on companies themselves?

- AI Now argues that much like the financial sector, AI industries will also have unaccounted or unimaginable impacts on the daily lives of people. AI Now believes that the government needs to mandate that companies monitor and mitigate the damages their products cause themselves. This is valid since the companies understand their products and it's limitations best and also it's their responsibility to ensure no damage is caused by their products.

- What are the windows for action that are identified? Which do you personally think are the most effective or promising?

- The windows for action are listed as follows:

  1. Minimize the amount of data obtained and collected by big tech firms.
  2. Enforce prevention of monopolies and anti-competitive practices in every shape and form.
  3. Regulate Large scale AI Systems and models.
  4. Displace Audits as the primary response to AI harms and impose strong laws on the rules implemented in large scale AI systems.

  I believe that points #2 shows the most promise. I don't believe it's possible for governments to curb sharing of information between companies, regardless of how much they try. My understanding is that the market for data is vast and big tech and the government are heavily invested in it. Additionally, #4 paints a highly optimistic pisture of the AI landscape. I don't think it's possible to audit every single AI system and model. Thus I believe that #2 and #3 are the most promising. #3 requires a fairly in-depth knowledge of AI and with the amount of innovation in the field, it's almost possible for anyone, especially the government to keep up.

  In addition to the previous points, #2 is where the government and it's agencies can make the most impact through the least amount of work. Even though the government doesn't need to or doesn't understand AI much, legal entities and companies is something they can easily regulate.

- The executive summary contains this quote:

"These are only a handful of examples, and what they make clear is that there is nothing about artificial intelligence that is inevitable. Only once we stop seeing AI as synonymous with progress can we establish popular control over the trajectory of these technologies and meaningfully confront their serious social, economic, and political impacts—from exacerbating patterns of inequality in housing, credit, healthcare, and education to inhibiting workers’ ability to organize and incentivizing content production that is deleterious to young people’s mental and physical health."

Do you agree with this assessment? How might we rethink our relationship with AI and with technology in order to avoid these potential negative outcomes?

I completely agree with this assessment. Through the rapid progress of AI and it's ability to discover solutions, people have ended up somewhat idolizing AI and it's capabilities. Instead of considering it as an oracle, I sternly believe that AI should be considered a tool to be used to solve problems instead of seeing it as a solution to a problem. It's extremely important to understand it's limitations and influence it's progress in an ehtical manner and ensure it produces results void of pre-exsting human bias. Exhaustively testing AI in a sandboxed environment ensures that we minimize it's negative outcomes and control it's harmful impact on people's lives.

Now paste this last question into ChatGPT and include its response. How do you compare its output to your own?

```
I don't have personal opinions or beliefs, but I can provide an analysis
of the statement and offer some insights into the discussion.

The statement highlights the importance of recognizing that artificial
intelligence (AI) is not an inevitable force for progress. It points out
that while AI and related technologies hold great potential, they can also
have significant negative impacts on various aspects of society, including
exacerbating inequalities, limiting workers' rights, and contributing to
content that may harm mental and physical well-being.

It is crucial to acknowledge that AI and technology, like any tool, are
neutral in themselves; their impact on society depends on how they are
developed, deployed, and regulated. To address the potential negative
outcomes associated with AI and technology, we can consider the following
strategies:

    Ethical AI Development: Prioritize the development of AI technologies
with a strong ethical foundation. This includes principles like fairness,
transparency, accountability, and bias mitigation.

    Regulation and Governance: Establish clear and comprehensive
regulations that guide the development and use of AI. Regulations can
address issues like data privacy, algorithmic transparency, and the
responsible use of AI in various domains.

    Inclusive Design: Involve diverse voices and perspectives in the
design and development of AI systems to ensure they serve the needs and
interests of all stakeholders.

    Education and Awareness: Promote AI literacy and awareness among the
general public. Understanding the potential and limitations of AI can
empower individuals to make informed decisions about its use.

    Responsible Content Creation: Encourage platforms and content creators
to adhere to ethical standards and avoid producing content that may be
harmful or detrimental to individuals, particularly young people.

    Worker Protections: Implement policies and regulations that protect
workers' rights and well-being in the face of increased automation and AI
technology in the workplace.
```

```
    Access and Inclusivity: Ensure that AI and technology are accessible
to all, reducing the risk of exacerbating inequalities.

    Oversight and Accountability: Establish mechanisms for monitoring and
holding organizations and developers accountable for the impact of their
AI systems on society.

    Public Engagement: Engage the public in discussions about the
direction of AI and technology, allowing for democratic input and
decision-making on AI-related policies and applications.

The overarching idea is that AI and technology should be harnessed for the
benefit of society while mitigating potential negative consequences. This
requires a multi-faceted approach involving government, industry,
academia, and civil society to collectively address the ethical, social,
and economic challenges associated with AI. It's important to recognize
that the responsible and thoughtful development and deployment of AI
technologies can lead to positive outcomes while minimizing harm and
inequality.
```

I've pasted ChatGPT's response above. It seems a lot stronger in favor of AI and it's answer seems a lot more well-laid out. It focuses on the process to follow human approval more so than my own answer. It also seems to be a lot more optimistic about AI's future and it's ability to solve problems.