

English. Bokmål på side 6, nynorsk på side 9. I tilfelle du er usikker på betydningen av noen av begrepene (noen er ikke så lette å oversette), se på den engelske versjonen. **Answers in red text. If the student gets portions of the answer correct, reward accordingly.**

## Problem 1

20 points total, 1 point each subtask. Answer each question with *True* or *False*.

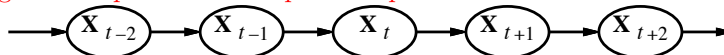
- (a) The course is about designing rational agents. **True**
- (b) Weak AI defines a machine that can act intelligently. **True.**
- (c) The Turing test is a good robot performance measure. **False, it is limited to text environments only. Even then, arguably not a good indicator of machine intelligence.**
- (d) Strong AI is the same as machine consciousness. **True.**
- (e) When you throw a dice twice, the probability of getting 6 at least once is  $\frac{1}{3}$ . **False, it is  $1 - (\frac{5}{6})^2 = \frac{11}{36}$**
- (f) Partial observability is not a problem in most cases. **False.**
- (g) A Boolean random variable is the best way to model temperature. **False, continuous variables is a much better choice.**
- (h) Conditional probability is the same as prior probability. **False, it is posterior probability.**
- (i) Structuring causal knowledge into belief networks is intuitive and helpful for humans. **True.**
- (j) Independence between variables is common in most applications. **False, it is extremely rare. Not to be confused with conditional independence.**
- (k) A rational agent does not have to always choose the action that maximizes expected utility. **False, a rational agent must always choose the action that maximizes expected utility.**
- (l) Humans are always irrational. **False.**
- (m) Stochastic dominance can often be determined with qualitative reasoning. **True.**
- (n) You add action nodes, utility nodes and policy nodes to a belief network in order to do rational decision making. **False, you only need an action and a utility node.**
- (o) Case-based reasoning is a form of supervised learning. **True.**

- (p) Neural networks are typically trained with the backpropagation algorithm. **True.**
- (q) Neural networks accurately represent all the properties of real neurons. **False, they are a gross oversimplification.**
- (r) A perceptron can be used to represent any function. **False, not those that are not linearly separable.**
- (s) Backpropagation is the same as gradient descent. **True.**
- (t) The real world is a Partially Observable Markov Decision Process. **True, with unknown transition and observation models.**

## Problem 2

20 points total, 4 points each subtask.

- (a) What is the Chinese Room argument, and how does it relate to weak/strong AI? **The Chinese Room argument is a person who follows a rule book (in English) that dictates how to respond based on the Chinese input. He does not understand the input nor the output, in other words it is not intelligent, nor is it thinking. This is neither weak nor strong AI.**
- (b) Explain what a first-order Markov process is. Give a simple example with a figure. **Simply put, the future is conditionally independent of the past given the present. A simple example can be seen below.**



- (c) Explain what filtering, prediction, smoothing and most likely explanation is.

**Filtering:**  $P(X_t|e_{1:t})$ . This is the *belief state* – input to the decision process of a rational agent.

**Prediction:**  $P(X_{t+k}|e_{1:t})$  for  $k > 0$ . Evaluation of possible action sequences; like filtering without the evidence

**Smoothing:**  $P(X_k|e_{1:t})$  for  $0 \leq k < t$ . Better estimate of *past* states – Essential for learning

**Most likely explanation:**  $\arg \max_{x_{1:t}} P(x_{1:t}|e_{1:t})$ . Speech recognition, decoding with a noisy channel

Note: the student is not required to write down the expressions as above, explaining with text is sufficient. However, if the answer is very short, this can be compensated for with the inclusion of the expressions.

- (d) How do you choose the “split attribute” when building a decision tree? A good attribute splits the examples into subsets that are all positive/negative. In order to do this, use a heuristic that minimizes the remaining information needed to classify a new sample. This is the information gain.
- (e) Briefly describe the most important properties of deep learning. Deep learning neural networks are characterized by many hidden layers, with specific properties at each layer (e.g. different transfer functions), where each layer represent some abstraction of the input signal. Feature engineering is not performed, instead raw data is used.

### Problem 3

20 points total, 10 points each subtask.

- (a) Describe what is case-based reasoning, and highlight its strengths and weaknesses. Here the student should mention the entire CBR cycle (retrieve - reuse - revise - retain), and elaborate upon this. It is well suited to many domains where it is difficult to formulate a mathematical model. The case base can be said to be similar to human memory. Disadvantages include that cases often don't include deeper knowledge of the domain, and it is difficult to determine good criteria for indexing and matching cases.
- (b) Describe reinforcement learning, and how it differs from the other learning algorithms in the course. A way to train an agent in unknown environments where there are intermittent delayed rewards. The rewards can be on varying time scales. Agent learns on its own. An important question is how to balance exploration vs exploitation. Reinforcement learning uses Q-learning, which is model free, to calculate expected discounted cumulative rewards. It is different from neural networks and decision trees which require labeled training data for each “timestep”. This makes reinforcement learning better suited for dynamic and unknown environments, such as finding the path in a maze without a map.

### Problem 4

20 points total, 10 points each subtask.

You are given the task of improving the music suggestion system on Spotify. The system should to predict whether the listener will like the song, based on whether there are good reviews of the song and whether his/her friends like it. Assume we have two songs in the system:  $s_1$  and  $s_2$ , each with uniform priors on their possibilities. The system already knows that  $s_1$  has a probability of good reviews of 0.8,  $s_2$  has 0.3. Similarly, his/her friends like  $s_1$  with a probability of 0.7, and  $s_2$  with a probability of 0.2. The system also knows that if

a song has a good review, and his/her friends like it, the user will like the song with a probability of 0.9. If there are good reviews, but none of his/her friends like it, the probability of liking the song is 0.4. If there are no good reviews, but his/her friends like the song, the probability of liking the song is 0.8. If there are no good reviews and no friends like it, the probability of liking the song is 0.1.

- (a) Model the domain as a Bayesian network. Make the model as simple as possible. Show the graphical structure and the conditional probability tables.

Graphical structure:

Song  $\rightarrow$  Good reviews, Friends like it  
 Good reviews  $\rightarrow$  Like  
 Friends like it  $\rightarrow$  Like

Tables:

Song  
 s1 | 0.5  
 s2 | 0.5

Good reviews  
 s1 | 0.8  
 s2 | 0.3

Your friends like it  
 s1 | 0.7  
 s2 | 0.2

From now on “Good reviews” is abbreviated “G” and “Your friends like it” as “Y”.

Like  
 G, Y | 0.9  
 G,  $\neg$ Y | 0.4  
 $\neg$ G, Y | 0.8  
 $\neg$ G,  $\neg$ Y | 0.1

- (b) Calculate the probability that the listener will like s1:  
 $P(\text{Like}=\text{true} \mid \text{Song}=\text{s1})$  This is inference by enumeration. Credit will also be given if the student cannot calculate properly, but can explain how to do it. To compute this, we must sum over the hidden variables.

$$P(\text{Like}=\text{true} \mid \text{Song}=\text{s1}) = \sum_G \sum_Y P(\text{Like}=\text{true} \mid G, Y) * P(G \mid \text{Song}=\text{s1}) * P(Y \mid \text{Song}=\text{s1})$$

Calculations:

$$G, Y : 0.9 * 0.8 * 0.7$$

$$G, \neg Y : 0.4 * 0.8 * 0.3$$

$$\neg G, Y : 0.8 * 0.2 * 0.7$$

$$\neg G, \neg Y : 0.1 * 0.2 * 0.3$$

$$= 0.9 * 0.8 * 0.7 + 0.4 * 0.8 * 0.3 + 0.8 * 0.2 * 0.7 + 0.1 * 0.2 * 0.3 = 0.718$$

## Oppgave 1

20 poeng total, 1 poeng for hver deloppgave. Svar hvert spørsmål med *True* eller *False*.

- (a) Faget handler om å designe rasjonelle agenter.
- (b) Svak AI defineres som en maskin som kan gjøre ting på en intelligent måte.
- (c) Turing-testen er en god måte for å måle en robots ytelse.
- (d) Sterk AI er det samme som maskin-bevissthet.
- (e) Når du kaster en terning to ganger, er sannsynligheten for å få 6 minst én gang  $\frac{1}{3}$ .
- (f) Partiell observerbarhet er ikke et problem i de fleste tilfeller.
- (g) En boolsk tilfeldig variabel er den beste måten å modellere temperatur på.
- (h) Kondisjonale sannsynligheter er det samme som prior sannsynlighet.
- (i) Å strukturere kausal kunnskap i tros-nettverk er intuitivt og hjelpsomt for mennesker.
- (j) Uavhengighet mellom variabler er vanlig i de fleste applikasjoner.
- (k) En rasjonell agent må ikke alltid velge handlingen som maksimerer den forventede nytten.
- (l) Mennesker er alltid irrasjonelle.
- (m) Stokastisk dominans kan ofte bli bestemt med kvalitativ resonnering.
- (n) Du legger til aksjons-noder, nytte-noder og politikk-noder til et tros-nettverk for å gjøre rasjonell beslutningstaking.
- (o) Case-basert resonnering er en form for overvåket læring.
- (p) Nevrale nettverk trenes typisk med tilbakepropageringsalgoritmen.
- (q) Nevrale nettverk representerer på en god måte alle egenskapene ved ekte nevroner.
- (r) En perceptron kan brukes til å representere hvilken som helst funksjon.
- (s) Tilbakepropagering er det samme som gradient nedstigning.
- (t) Den virkelige verden er en delvis observerbar Markov beslutnings-prosess.

## Oppgave 2

20 poeng totalt, 4 poeng på hver deloppgave.

- (a) Hva er det kinesiske rom-argumentet, og hvordan relateres det til svak/sterk AI?
- (b) Forklar hva en første-ordens Markov prosess er. Gi et enkelt eksempel med en figur.
- (c) Forklar hva filtrering, prediksjon, glatting og mest sannsynlige forklaring er for noe.
- (d) Hvordan velger du “splitt-attributtet” når du bygger et beslutningstre?
- (e) Beskriv kort de viktigste egenskapene ved deep learning.

## Oppgave 3

20 poeng totalt, 10 poeng hver deloppgave.

- (a) Beskriv case-basert resonnering, og vektlegg dets styrker og svakheter.
- (b) Beskrive forsterkende læring, og hvordan det skiller seg fra de andre læringsalgoritmene i faget.

## Oppgave 4

20 poeng totalt, 10 poeng hver deloppgave.

Du har fått i oppgave å forbedre Spotify sitt system for å foreslå musikk. Systemet skal predikere hvorvidt lytteren vil like en sang, basert på hvorvidt det er gode anmeldelser av sangen og hvorvidt hans/hennes venner liker sangen. Anta at vi har to sanger i systemet,  $s_1$  og  $s_2$ . Systemet vet allerede at  $s_1$  har en sannsynlighet for gode anmeldelser som er 0.8,  $s_2$  har 0.3. Hans/hennes venner liker  $s_1$  med sannsynlighet på 0.7, og  $s_2$  med en sannsynlighet på 0.2. Systemet vet også at hvis en sang har en god anmeldelse og vennene liker sangen, så vil lytteren like sangen med en sannsynlighet på 0.9. Hvis det er gode anmeldelser, men ingen venner som liker sangen, er sannsynligheten for å like sangen 0.4. Hvis det er ingen gode anmeldelser, men vennene liker sangen, er sannsynligheten for å like sangen 0.8. Hvis det er ingen gode anmeldelser og ingen venner liker sangen, er det en sannsynlighet på 0.1 for at lytteren vil like sangen.

- (a) Modellér domenet som et Bayesiansk nettverk. Gjør modellen så enkel som mulig. Vis den grafiske strukturen og de kondisjonelle sannsynlighetstabellene.

- (b) Regn ut sannsynligheten at lytteren vil like  $s_1$ :  
 $P(\text{Like}=\text{true} \mid \text{Song}=\text{s}_1)$ .



## Oppgave 1

20 poeng total, 1 poeng for kvar deloppgåve. Svar kvart eit spørsmål med *True* eller *False*.

- (a) Faget handlar om å designe rasjonelle agentar.
- (b) Svak AI definerast som ein maskin som kan gjera ting på ein intelligent måte.
- (c) Turing-testen er ein god måte for å måle ein robots ytelse.
- (d) Sterk AI er det samme som maskin-bevisstheit.
- (e) Når du kaster ein terning to gonger, er sannsynligheita for å få 6 minst éin gong  $\frac{1}{3}$ .
- (f) Partiell observerbarheit er ikkje eit problem i de fleste tilfelle.
- (g) En boolsk tilfeldig variabel er den beste måten å modellere temperatur på.
- (h) Kondisjonale sannsynlighet er det samme som prior sannsynlighet.
- (i) Å strukturere kausal kunnskap i tros-nettverk er intuitivt og hjelpesamt for menneskje.
- (j) Uavhengigheit mellom variablar er vanlig i dei fleste applikasjonar.
- (k) Ein rasjonell agent må ikkje alltid velja den handlinga som maksimerar den forventede nytta.
- (l) Menneskje er alltid irrasjonelle.
- (m) Stokastisk dominans kan ofte bli bestemt med kvalitativ resonnering.
- (n) Du leggjer til aksjons-noder, nytte-noder og politikk-noder til et tros-nettverk for å gjera rasjonell beslutningstaking.
- (o) Case-basert resonnering er ein form for overvåka læring.
- (p) Nevrale nettverk trenes typisk med tilbakepropageringsalgoritmen.
- (q) Nevrale nettverk representerer på ein god måte alle egenskapane ved ekte nevronar.
- (r) En perceptron kan brukast til å representere ein vikarleg funksjon.
- (s) Tilbakepropagering er det samme som gradient nedstigning.
- (t) Den virkelige verda er ein delvis observerbar Markov beslutnings-prosess.

## Oppg ve 2

20 poeng totalt, 4 poeng kvar deloppg ve.

- Kva er det kinesiske rom-argumentet, og korleis relateres det til svak/sterk AI?
- Forklar kva en f rste-ordens Markov prosess er. Gi et enkelt eksempel med ein figur.
- Forklar kva filtrering, prediksjon, glatting og mest sannsynlige forklaring er for noko.
- Korleis velgjer du "splitt-attributtet" n r du byggjar eit beslutningstre?
- Beskriv kort dei viktigste egenskapene ved deep learning.

## Oppg ve 3

20 poeng totalt, 10 poeng kvar deloppg ve.

- Beskriv case-basert resonnering, og vektlegg dets styrkar og svakhetar.
- Beskrive forsterkande l ring, og korleis det skil seg fra de andre l ringsalgoritmane i faget.

## Oppg ve 4

20 poeng totalt, 10 poeng kvar deloppg ve.

Du har f tt i oppgave   forbetra Spotify sitt system for   foresl  musikk. Systemet skal predikera om lytteren vil like ein sang, basert p  om det er gode anmeldelser av sangen og om hans/hennar venner liker sangen. Anta at vi har to sangar i systemet,  $s_1$  og  $s_2$ . Systemet vet allereie at  $s_1$  har ein sannsynlighet for gode anmeldelser som er 0.8,  $s_2$  har 0.3. Hans/hennar venner liker  $s_1$  med sannsynlighet p  0.7, og  $s_2$  med ein sannsynlighet p  0.2. Systemet veit ogs  at hvis ein sang har ein god anmeldelse og vennane liker sangen, s  vil lyttaren like sangen med ein sannsynlighet p  0.9. Viss det er gode anmeldelsar, men ingen vennar som liker sangen, er sannsynligheta for   like sangen 0.4. Viss det er ingen gode anmeldelser, men vennene liker sangen, er sannsynligheta for   like sangen 0.8. Viss det er ingen gode anmeldelser og ingen venner liker sangen, er det ein sannsynlighet p  0.1 for at lytteren vil like sangen.

- Modell r domenet som eit Bayesianisk nettverk. Gjer modellen s  enkel som mogleg. Vis den grafiske strukturen og dei kondisjonelle sannsynlighetstabellane.
- Regn ut sannsynligheta for at lyttaren vil like  $s_1$ :  
 $P(\text{Like}=\text{true} \mid \text{Song}=s_1)$ .