TDT4171 - Artificial Intelligence Methods Final Exam Spring 2024

Introduction

- The exam contains 9 tasks with several subtasks in each task.
- Each subtask is worth a certain number of % points, adding up to the number of points specified for the task and 100% for the complete exam.
- If the description of the task is ambiguous, specify your assumptions and solve the task according to these assumptions.
- To get maximum score, the answer should be correct, complete, well-justified and precise. Partial points will be given for answers that are somewhat lacking in content or quality.
- In several tasks, we specified a range for how many sentences should be used for the answer. This is a recommended number of sentences to give you an idea about the expected level of detail. Minor deviations from this range are acceptable. If your answer is significantly longer, consider summarizing the main points in fewer sentences. If your answer is shorter, consider expanding it a bit.

1 [6%] AI foundation

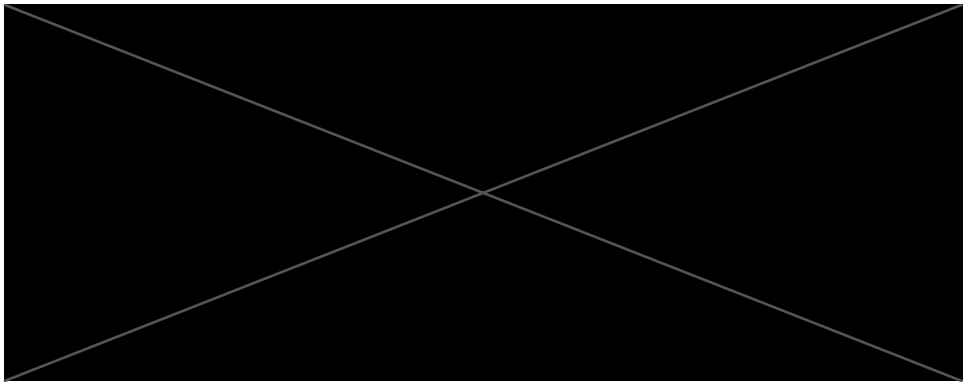
1.1 [3%]

Which of these statements are True or False? +0.5% for correct and -0.5% for incorrect, 0% points for no answer.

1. The main goal of this course is to understand what intelligence is.
2. By definition, Strong AI outperforms Weak AI in all tasks.
3. Weak AI can pass a Turing test.
4. Acting rationally means acting like humans do.
5. An AI agent with faulty sensors can still act rationally.
6. To be rational, an agent must be able to compute utilities and probabilities.

1.2 [3%]

List any five ethical concerns / misuses of AI.



2 [10%] Uncertainty and Bayesian networks

2.1 [3%]

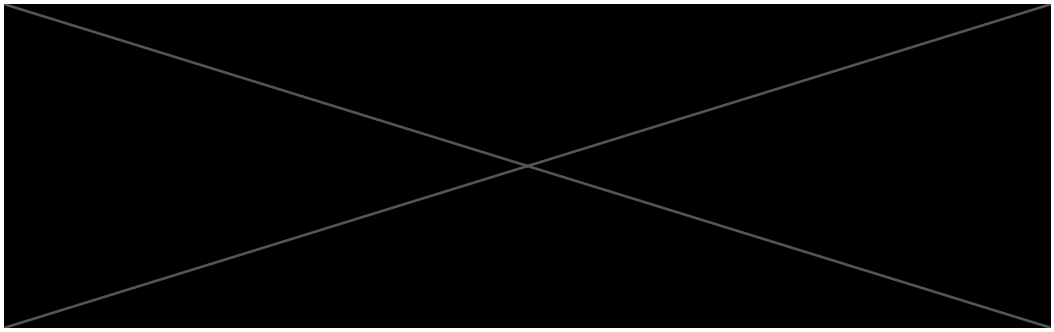
Given the following joint probability distribution:

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.13	.03	.07	.01
¬ cavity	.01	.06	.12	.57

Compute the following probabilities:

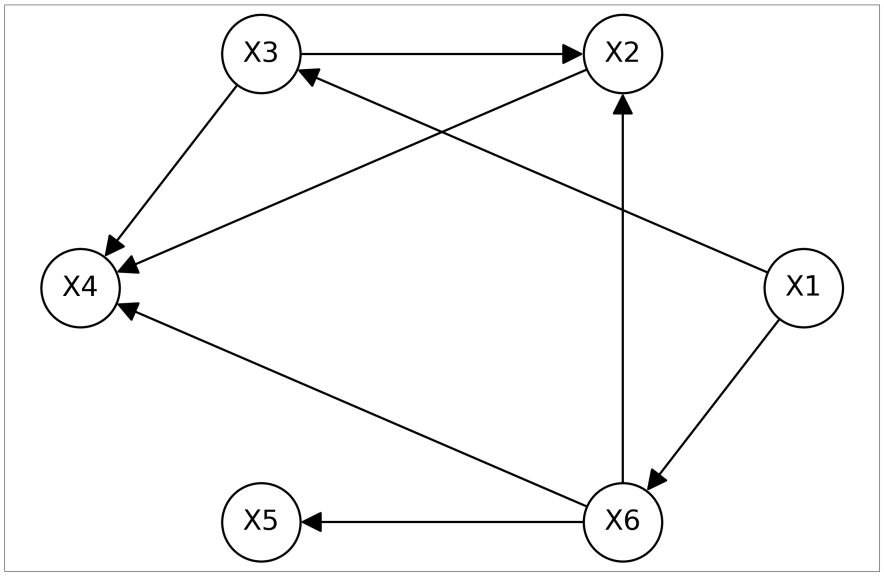
1. $P(toothache)$
2. $P(cavity \vee toothache)$
3. $P(\neg cavity \mid toothache)$

Round your answers to two decimals, e.g. 1.236 should be rounded to 1.24



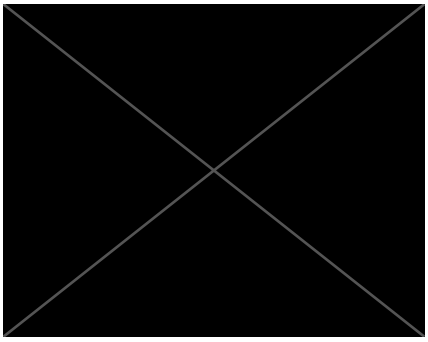
2.2 [**7%**]

Given the following Bayesian network:



Answer these Yes/No questions. +1% for correct and -1% for incorrect, 0% points for no answer. +2% for all correct.

1. Is X1 independent of X4?
2. Is X3 independent of X6 given X2?
3. Is X1 independent of X5 given {X4, X6}?
4. Is X1 independent of X2 given {X4, X5, X6}?
5. Is X3 independent of X5 given {X1, X2, X4, X6}?



3 [16%] Decision networks

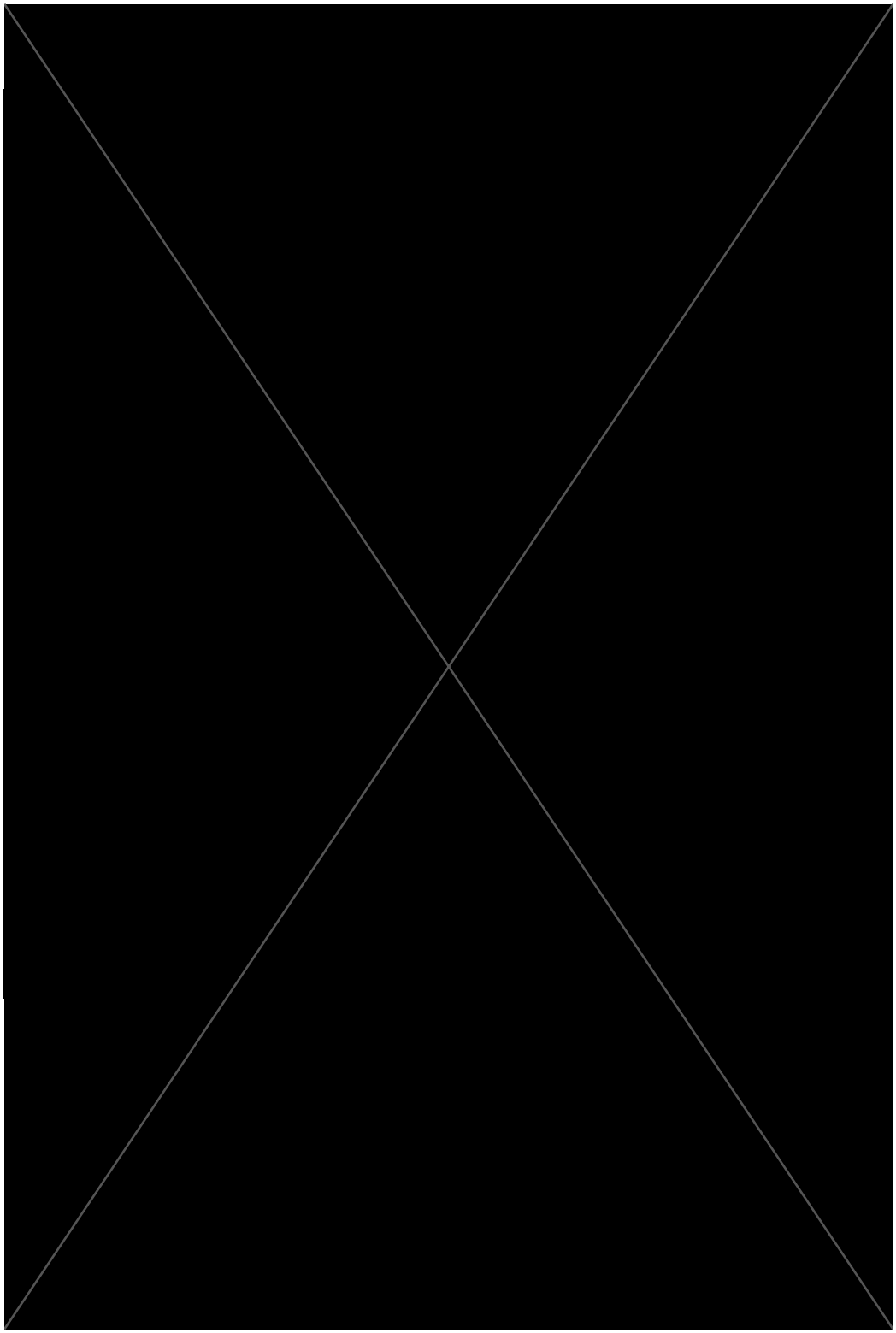
3.1 [5%]

Consider a scenario where a doctor is diagnosing a patient with some symptoms. The doctor needs to make two decisions: (1) what kind of test to perform and (2) what treatment to prescribe.

Construct the decision network for this scenario. Include nodes for disease, symptoms, test, test cost, test results, treatment, treatment outcome and utility. Use ovals for chance nodes, rectangles for decision nodes and diamonds for utility nodes.

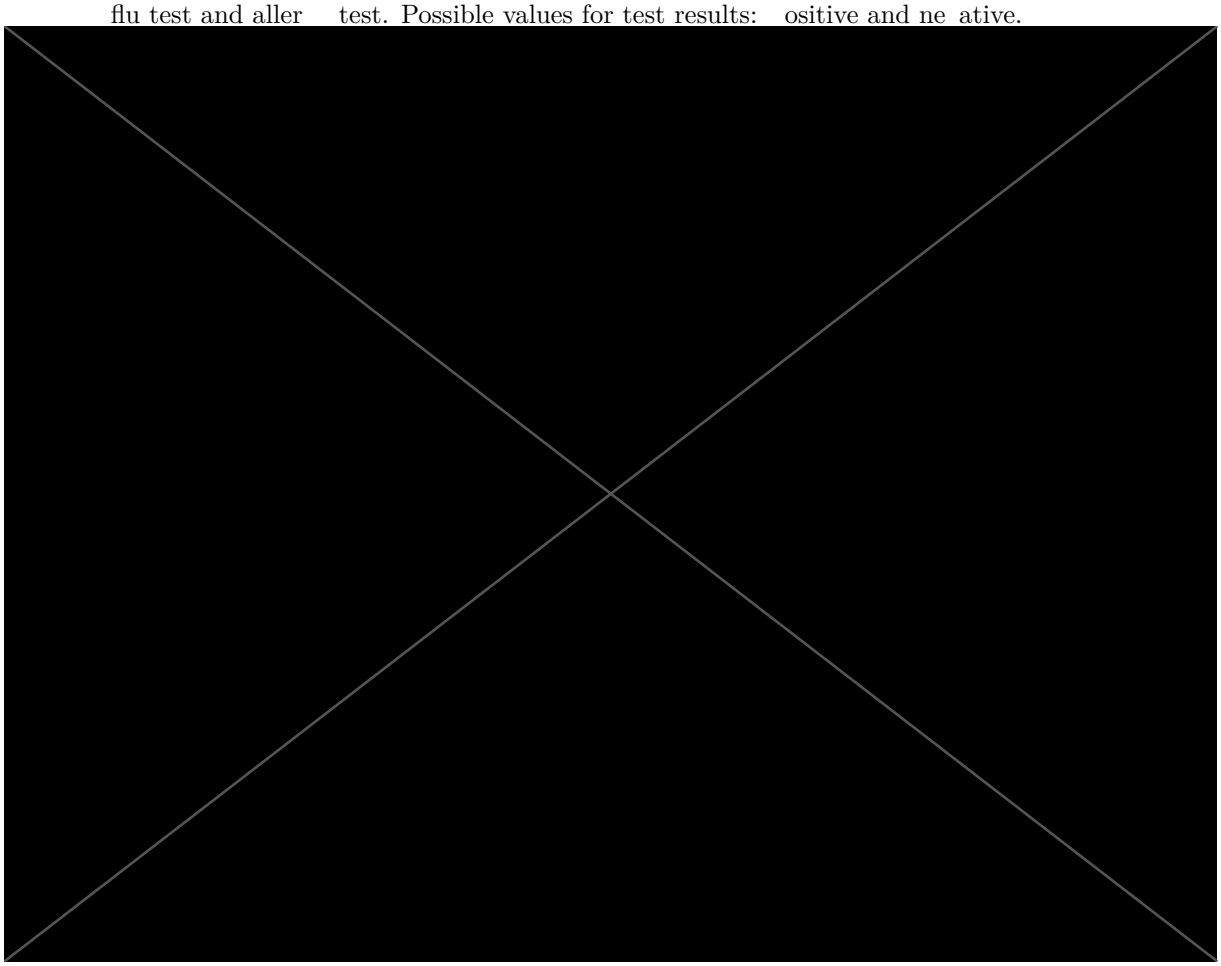
Draw the structure (the directed acyclic graph) of this model on **paper**. You are not asked to provide conditional probability tables. However, make your model so that it would be as easy as possible for a domain expert to provide the required probabilities.

Solution



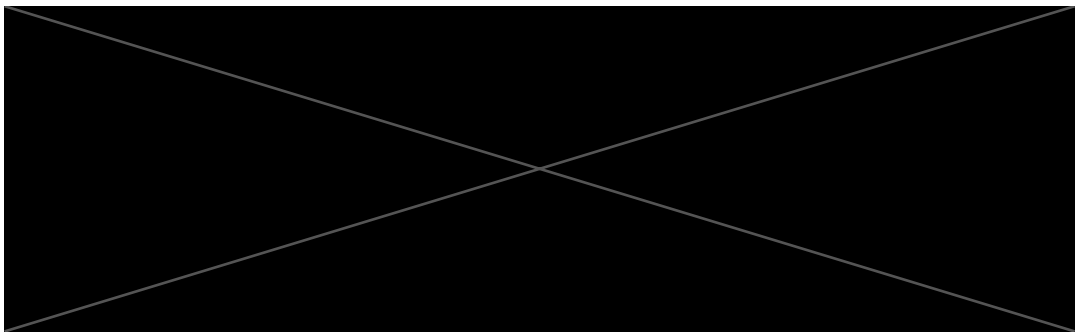
3.2 [4%]

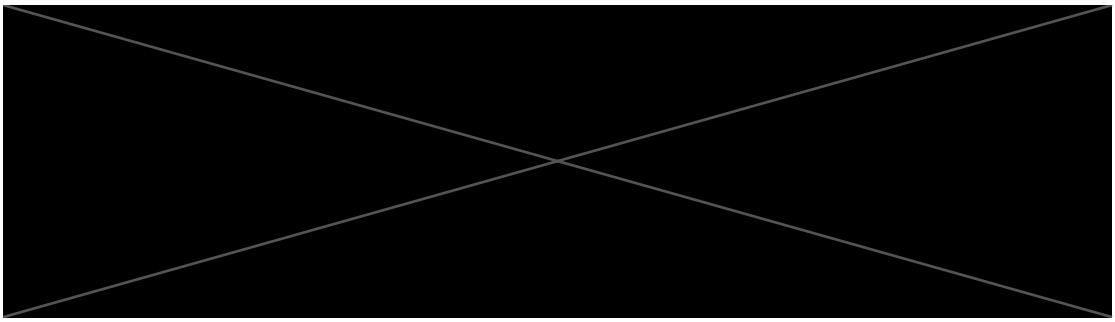
Given 3.1, specify the conditional probability table for the node representing test result. Possible values for disease: flu and allergy. Possible values for test: flu test and aller test. Possible values for test results: ositive and ne ative.



3.3 [3%]

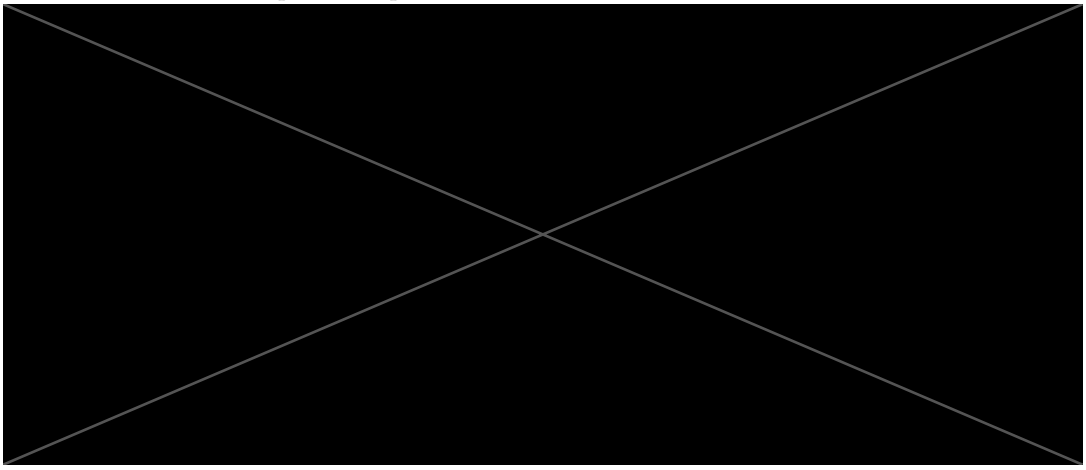
Design utility for the network in 3.1 The utility should be based on outcome and test cost. You may use an additive value function. Possible values for the outcome: No symptoms, fewer symptoms, same symptoms. Possible values for the test cost: high, moderate, low.





3.4 [4%]

Specify the general algorithm (briefly describe steps) for evaluating decision networks, i.e. how they are used for making decisions. You don't need to explain how to calculate posterior probabilities.



4 [20%] Probabilistic reasoning over time - Hidden Markov model

One of the challenges with solar energy production in the Nordic countries is snow covering solar panels. Solar panels can still produce some energy if a snow layer is thin or only part of the panel is covered. There are no sensors that can measure it directly, so we have to infer whether the panel is covered based on measured production.

Consider the following probabilities:

- Given that the panel is covered with snow, the probability that a solar panel produces energy is 0.1.
- Given that the panel is **not** covered with snow, the probability that a solar panel produces energy is 0.7.

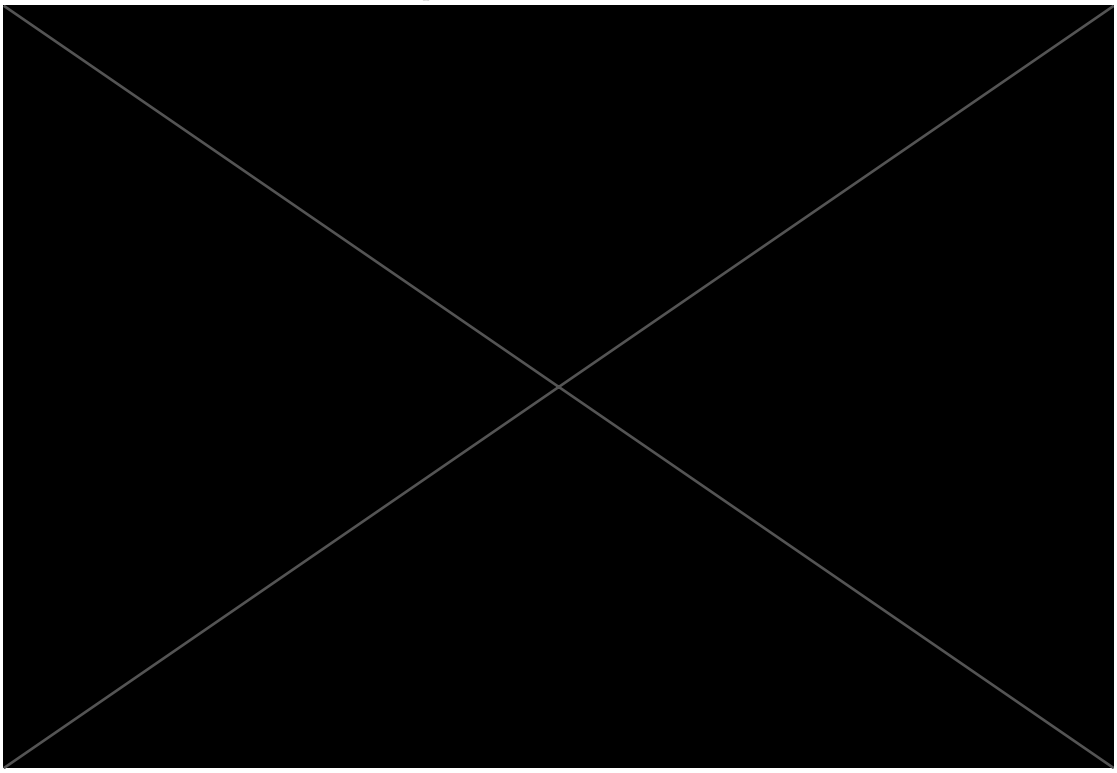
- Given that the panel is covered with snow during the current hour, the probability that the panel will be covered with snow for the next hour is 0.9.
- Given that the panel is **not** covered with snow during the current hour, the probability that it will **not** be covered with snow for the next hour is 0.8.
- The prior probability for the panel being covered with snow is 0.3.

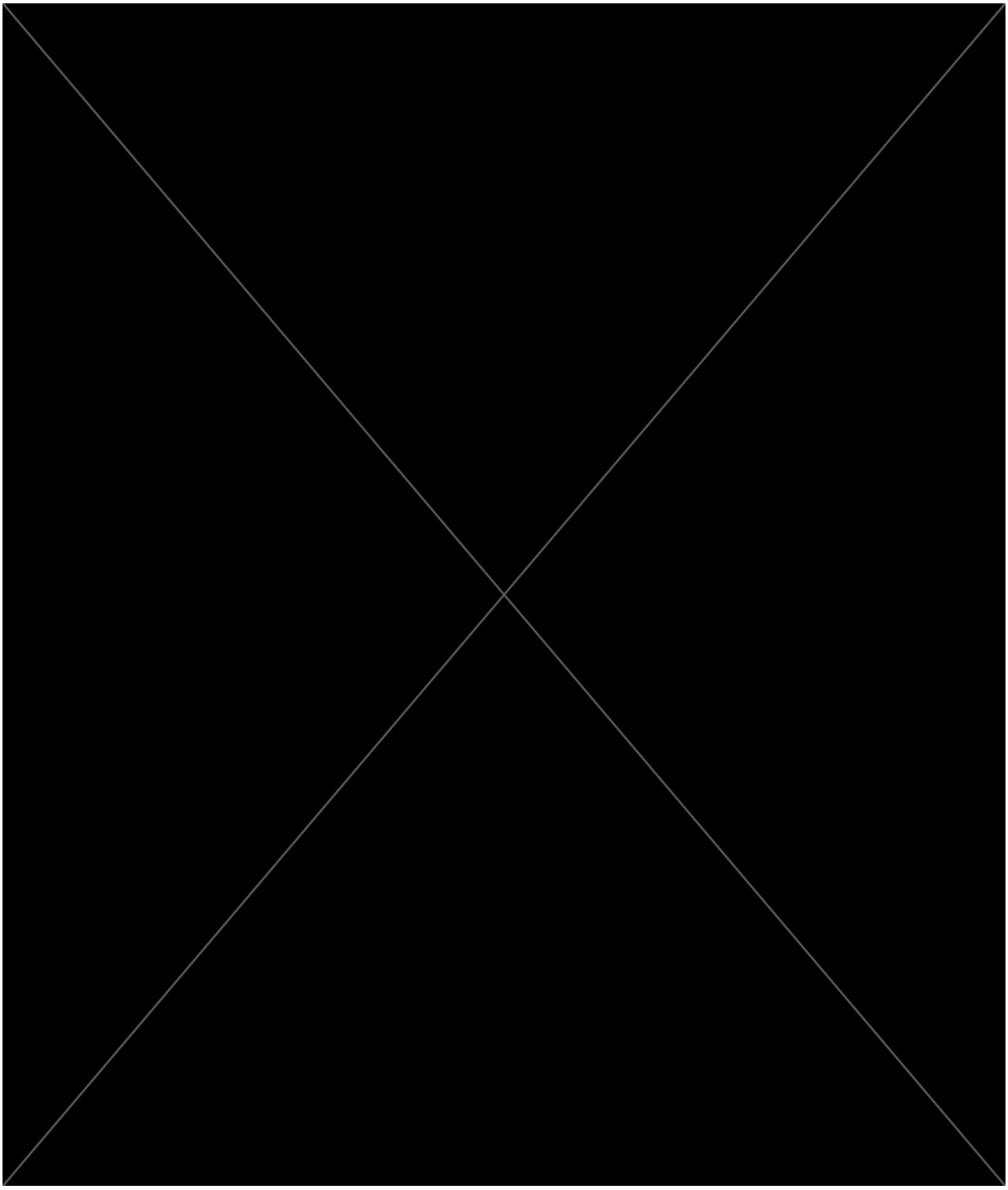
4.1 [8%]

Specify a hidden Markov model representing the problem above. You need to specify the structure and the probability tables. The structure should include steps (hours) 0, 1 and 2. Draw the model and tables on **paper**.

Use the following notation:

- S_t - random variable for the panel covered with snow at step t
- s_t - value for S_t at step t
- E_t - random variable for the panel producing energy at step t
- e_t - value for E_t at step t



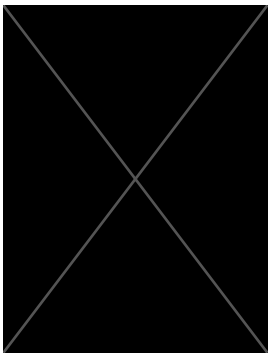


4.2 [3%]

Given the hidden Markov model from 4.1, are these statements true or false? +0.5% for correct and -0.5% for incorrect, 0% points for no answer.

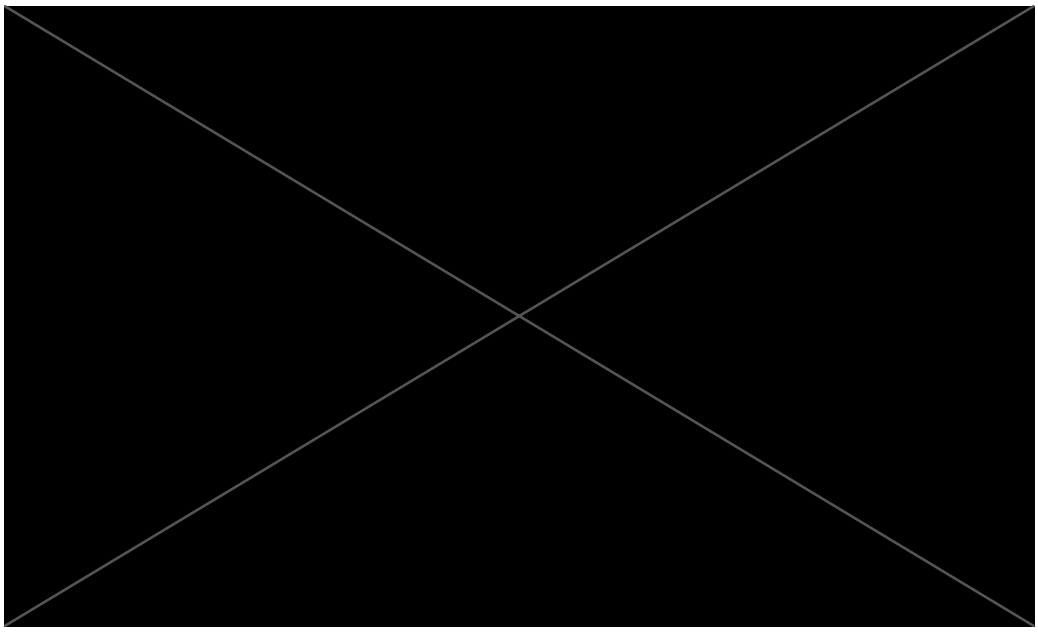
1. $P(S_t|E_t)$ is called the observation model.

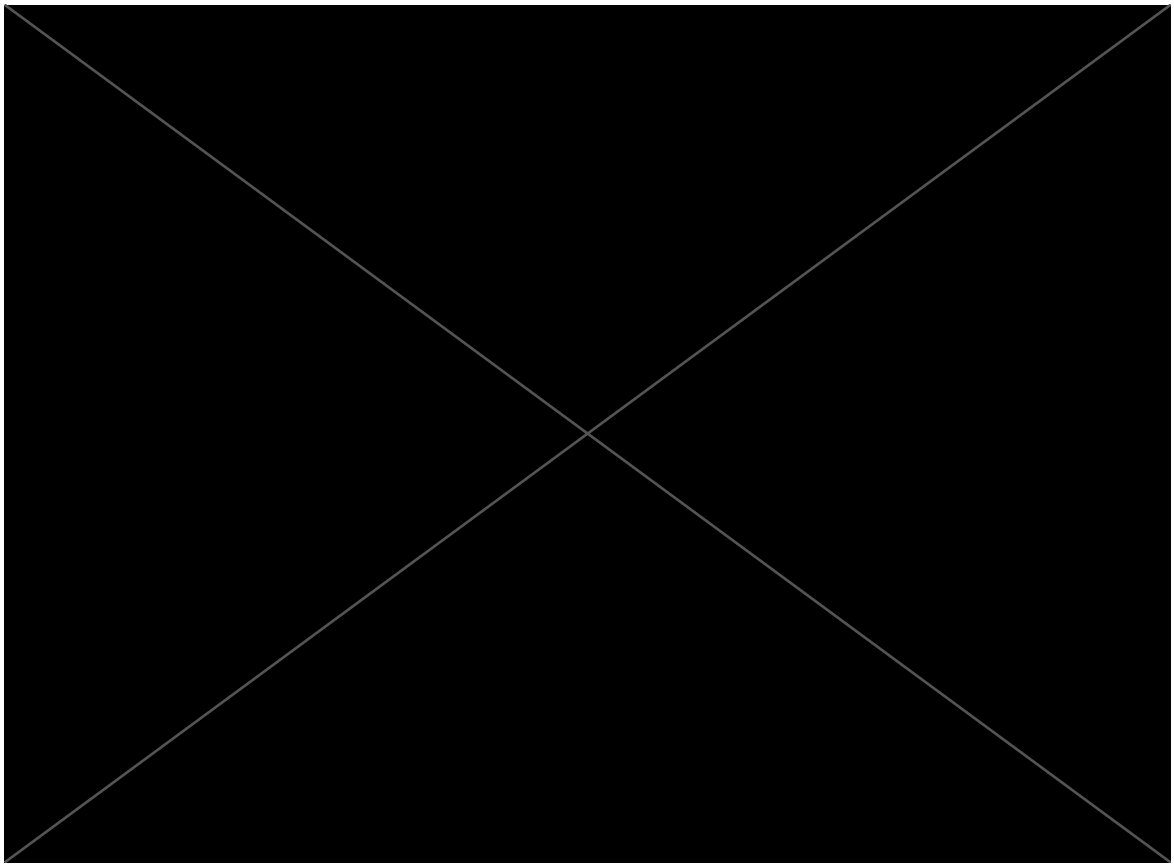
2. First-order Markov assumption can be expressed as $P(S_t|S_{0:t-1}) = P(S_t|S_{t-1})$
3. Filtering is the task of computing $P(S_k|e_{1:t})$ where t is the most recent time step and $0 \leq k < t$.
4. Prediction is the task of computing $P(S_{t+k}|e_{1:t})$ where t is the most recent time step and $k > 0$.
5. Smoothing is the task of computing $P(S_t|e_{1:t})$ where t is the most recent time step.
6. Most likely explanation is the task of computing $P(S_{1:t}|e_{1:t})$ where t is the most recent time step.



4.3 [9%]

Compute the probability of the panel being covered with snow at the second hour given that our measurements show energy production at hour 1 and no production at hour 2. Show your computations step by step for steps 0, 1 and 2. Show computations in symbolic form (variables, probabilities and distributions) before replacing them with numbers. In you computations, round intermediate and final results to two decimals.





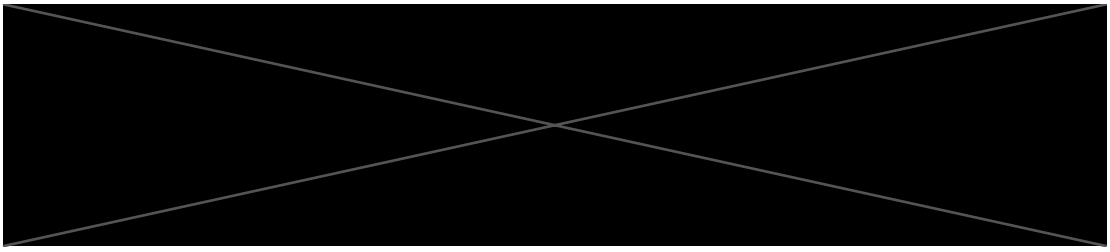
5 [10%] Artificial neural networks - Gradient descent

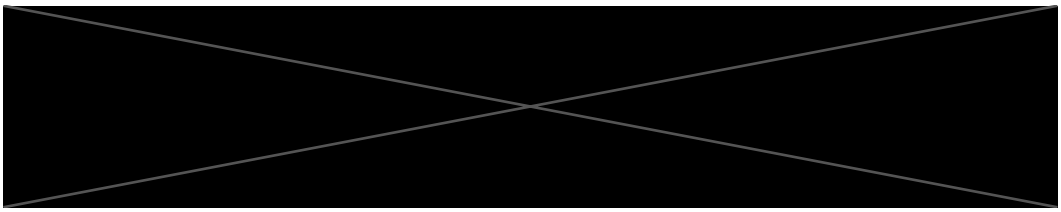
Given the Gradient Descent algorithm for the perceptron:

1. Initialize each w_i to some small random value
2. Until the termination condition is met:
    1. Initialize: $\Delta w_i \leftarrow 0$
    2. For each $\langle \vec{x}, t \rangle$ in D:
        - Input $\vec{x}$ to the unit and compute the output o
        - For each w_i: $\Delta w_i \leftarrow \Delta w_i - \eta \cdot 2(-x_i(t-o))$
    3. For each w_i: $w_i \leftarrow w_i + \Delta w_i$

5.1 [5%]

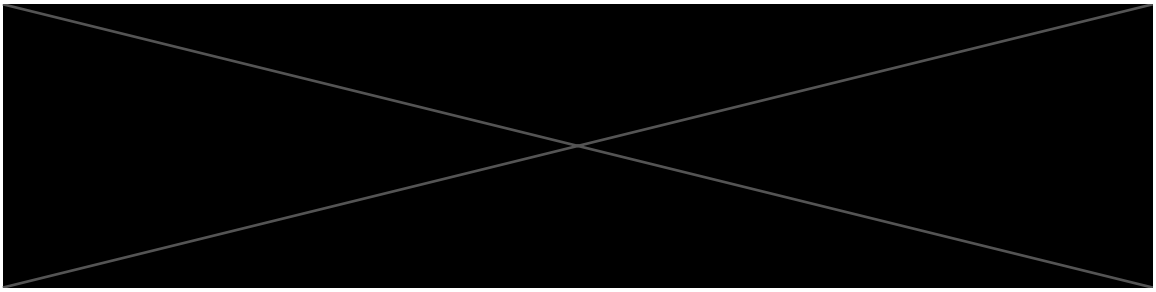
Explain what the following symbols mean and their role in this algorithm: D, $\vec{x}$, t, w, η. Answer with 1-2 sentences per symbol.





5.2 [5%]

Where is the gradient in this algorithm, and why does it make sense to use the gradient this way? Answer with 2-10 sentences.

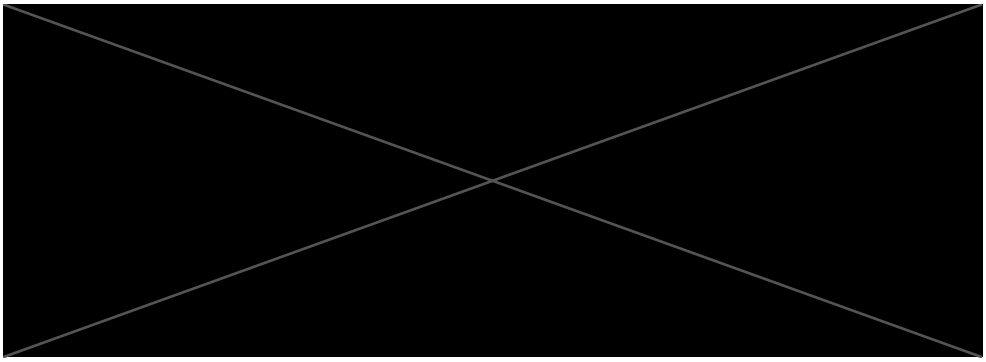


.

6 Deep learning [10%]

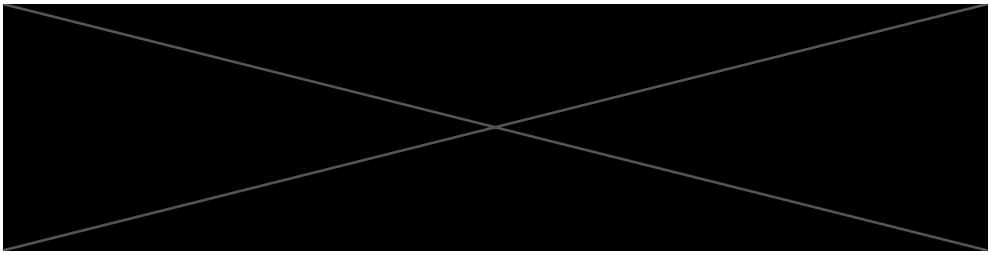
6.1 [2%]

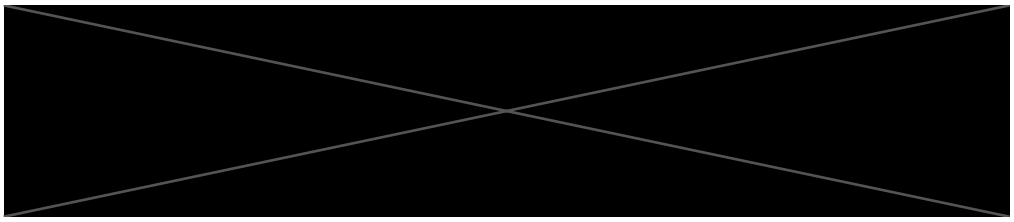
What is "deep" in deep learning, and why does it help? Answer with 3-10 sentences.



6.2 [4%]

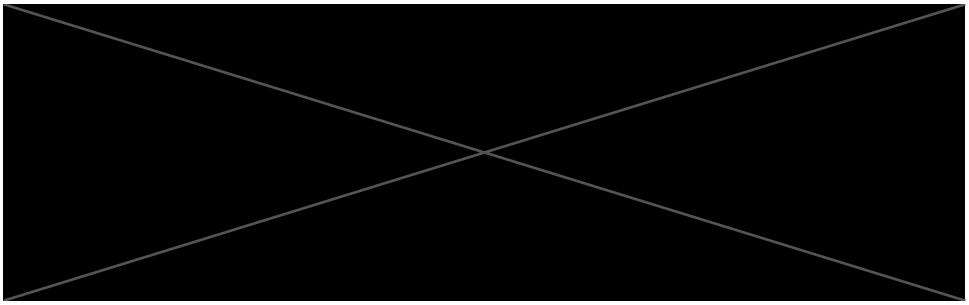
What is overfitting, and how can we avoid it in deep learning? Answer with 3-10 sentences.





6.3 [4%]

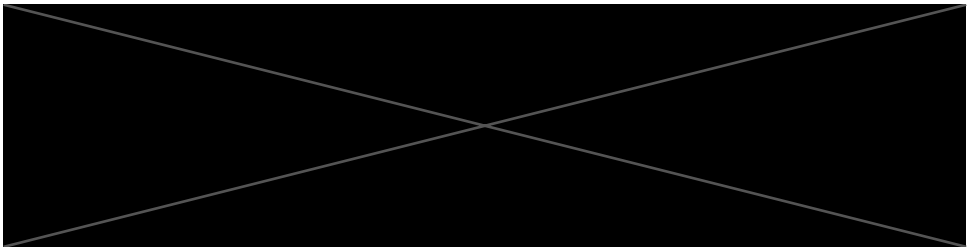
Explain why convolution networks work better for image data than a multilayer perceptron. Answer with 3-10 sentences.



7 [8%] Natural language processing

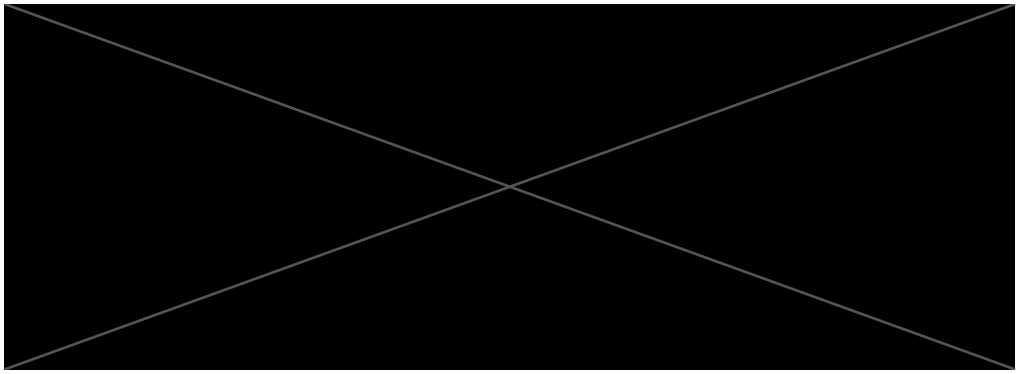
7.1 [4%]

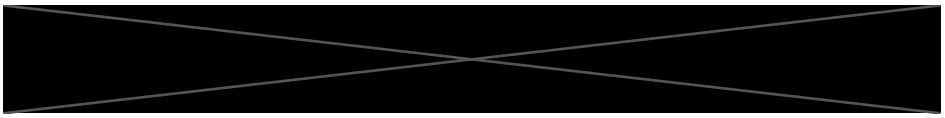
In natural language processing, what are word embeddings, and why are they better for many NLP tasks than one-hot encoding? Answer with 2-5 sentences.



7.2 [4%]

How would you use word embeddings to classify emails as spam, not spam? Here you only need to outline the general idea. Answer with 2-5 sentences.

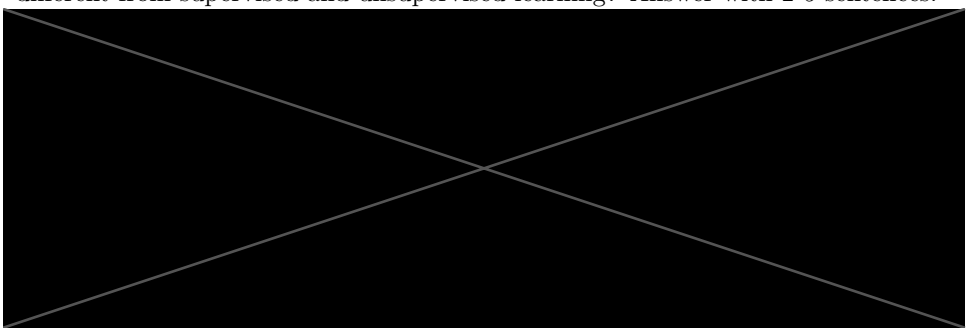




8 [10%] Reinforcement learning

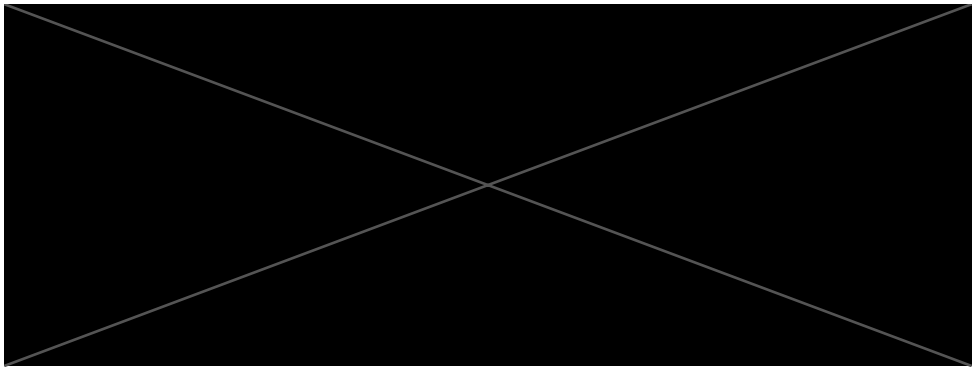
8.1 [2%]

What does "reinforcement" stand for in reinforcement learning? How is it different from supervised and unsupervised learning? Answer with 2-5 sentences.



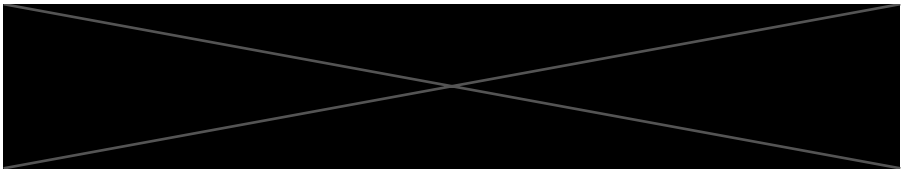
8.2 [4%]

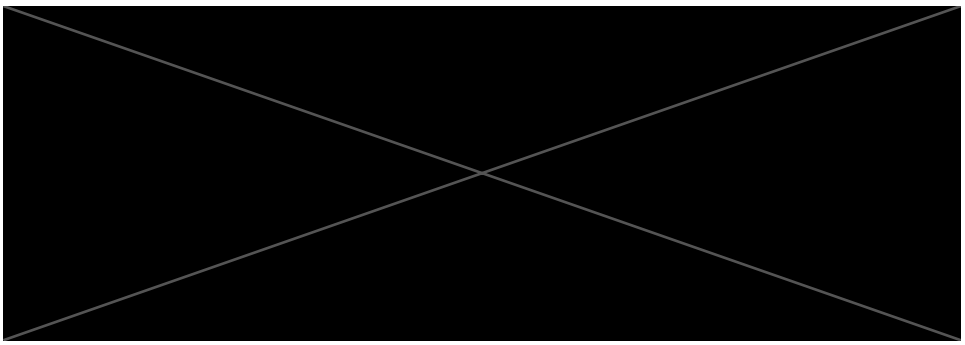
How is reinforcement learning (RL) different from solving a Markov decision process (MDP)? Answer with 2-5 sentences.



8.3 [4%]

Explain how deep learning is combined with reinforcement learning to play Flappy Bird. If you don't know Flappy Bird use Mario or some other simple arcade game as an example. You don't need to explain the complete architecture or details of the learning algorithm. Focus on how deep learning is integrated into reinforcement learning. Answer with 3-10 sentences.

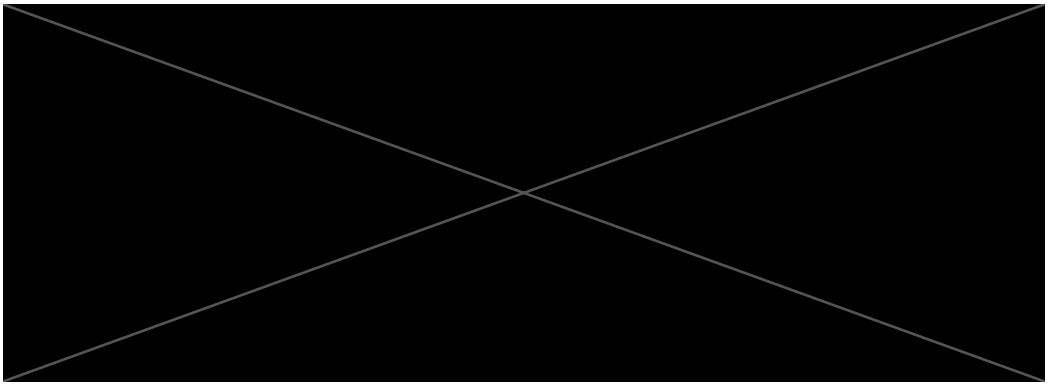




9 [10%] Case-based reasoning

9.1 [5%]

Explain each step in CBR cycle. 1-2 sentences per step.



9.2 [5%]

Find a problem/application where case-based reasoning would perform better than deep learning? Explain why it would be better for this problem. Answer with 3-10 sentences.