1 Grading Criteria

Students taking the course this year: only answer the 20 multiple-choice questions. The exam is pass/fail only.

Students who took the course in previous exam: please answer all the questions in order to get a 3/4/5 grade.

About the multiple choice questions:

- Each question gives a maximum of 5 points and a minimum of 0 points.
- In order to get the 5 points, you must select all correct alternatives and only the correct alternatives.
- Selecting a mixture of correct and incorrect statements will give you partial credits. Be aware that selecting, for example, 2 correct statements will give you a higher score than selecting 2 correct statements + 1 incorrect statement.
- Partial credit will vary depending on the number of correct alternatives.

Use the space below to write any assumptions, comments, etc regarding any of the questions.

Fill in your answer here

Maximum marks: 0

2 AI Basics

Which of the following statements are true:
Select one or more alternatives:

- [ ] Machine learning and AI are the same thing
- [ ] AI was originally defined as the simulation of human-like intelligence so that, for example, it would be impossible to differentiate between a machine and a human
- [ ] AI is concerned with building machines that can compute how to act effectively and safely in a wide variety of situations.
- [ ] Subareas of AI include blockchain, robotics, 3D printing, and virtual reality.
- [ ] Stochastic AI is concerned with problems in which either the environment or the actions include some form of randomness.

Maximum marks: 5

3 Search Basics

Which of the following statements are true?
Select one or more alternatives:

- [ ] Every problem can be represented as a search problem.
- [ ] If we choose tree search, DFS offers enormous memory advantages.
- [ ] The key difference between graph search and tree search is that the latter has a set of visited nodes.
- [ ] A search tree is recorded only when using tree search.
- [ ] The order of the nodes in the frontier does not matter.

Maximum marks: 5

4 Local Search

Which of the following statements are true?
Select one or more alternatives:

- [ ] Local beam search consists in running n random restarts concurrently instead of consecutively
- [ ] If we have m layers each with n nodes, dynamic programming reduces an optimal path finding problem from n^m to $m(n^2)$.
- [ ] Greedy hill climb is easy to implement and will most likely lead to the optimal solution
- [ ] Local beam search is an example of swarm intelligence
- [ ] The idea behind greedy hill climb is trying to find a good solution by always transitioning to the fittest neighbour, even though this is not always the case.

Maximum marks: 5

5 Copy of Planning and Scheduling I

Which of the following statements are true about planing problems?
Select one or more alternatives:

- [ ] PDDL problems can be solved using search algorithms, like best-first search
- [ ] In a PDDL problem, the goal state is unique
- [ ] Action specification is very important in order to avoid pointless searching
- [ ] Fluents include conjunctions, disjunctions, and conditionals.
- [ ] Actions are specified by a set of current preconditions and effects that can be immediate or at some point in the future.

Maximum marks: 5

6 Copy of Planning and Scheduling II

Which of the following statements about scheduling are true?
Select one or more alternatives:

- [ ] Scheduling is about when to do the required actions
- [ ] The goal is to provide a schedule that ends in the shortest amount of time
- [ ] Priority-based scheduling is always optimal
- [ ] Resource-free scheduling is a way to estimate a lower-bound of the end time of the tasks.
- [ ] It is not possible to obtain optimal solutions to scheduling problems, only good approximations

Maximum marks: 5

7 Copy of Competitive Search II

Which of the following statements are true about alpha-beta pruning?
Select one or more alternatives:

- [ ] Just like MiniMax, it can suffer from horizon effects but only when we limit the depth of the search
- [ ] Selecting initial moves that return a small [alpha,beta] interval of possible moves can help enormously to reduce the size of the search tree
- [ ] Compared to MiniMax, alpha-beta pruning returns an equally good but potentially different sequence of moves.
- [ ] It always returns the moves that leads to the best outcome in the shortest number of steps.
- [ ] Unlike MiniMax, alpha-beta pruning can speculate that an opponent might make a mistake

Maximum marks: 5

8 Copy of Competitive Search I

Which of the following statements are true about MiniMax?
Select one or more alternatives:

- [ ] MiniMax can be used for games like Chess and Go
- [ ] The move ordering is always fixed
- [ ] It can only be used for deterministic games with perfect information
- [ ] The number of nodes involved grows exponentially with depth
- [ ] A good move ordering is important and can greatly affect the performance of the algorithm

Maximum marks: 5

9 Copy of Shortest Path Problems

Which of the following statements are true about Shortest Path Problems?
Select one or more alternatives:

- [ ] Are applied to graphs that have one or more costs associated with the edges
- [ ] Keeping a set of visiting nodes always leads to a faster search, but sometimes we don't have enough memory
- [ ] If all edges have positive costs, they can be used to find the most efficient series of actions while minimizing some value
- [ ] The goal is to find a path from the origin to the goal, regardless of the cost
- [ ] Their associated graph is always acyclic.

Maximum marks: 5

10 Heuristic Search II

Which of the following statements are true about A*:
Select one or more alternatives:

- [ ] It is guaranteed to terminate and it is complete (i.e. it will always find a solution)
- [ ] The way in which we break ties when expanding a node in the frontier has no effect in the performance.
- [ ] Best-first search is just A* with a heuristic that always estimates the cost to the goal as 0
- [ ] A* always outperforms basic cumulative best-first search
- [ ] Optimistic heuristics are always monotonic

Maximum marks: 5

11 Copy of Local Search

In local search, how might we try to avoid getting stuck at local optima?
Select one or more alternatives:

- [ ] Sometimes transition to a worse/less fit node
- [ ] Re-running the search algorithm many times
- [ ] Use dynamic programming for path finding
- [ ] Always transition to a better/fitter node
- [ ] Randomly transition between nodes while keeping track of the best one seen so far

Maximum marks: 5

12 Basics

Which of the following statements are true:
Select one or more alternatives:

- [ ] A posterior distribution encodes domain knowledge before any data is observed
- [ ] P(A|B) is the conditional distribution of B given A
- [ ] A categorical variable is a random variable that can take only discrete and unordered values
- [ ] Statistical distributions have parameters, but statistical models don't.
- [ ] A distribution for a nominal variable assigns probabilities to each value of the variable

Maximum marks: 5

13 Markov Chains I

Replace with question text
Which of the following statements are true?

- [ ] The transition matrix gives the probabilities of observing different emissions given the state of the system
- [ ] In a graphical representation of a Markov chain, variables are represented by nodes, while dependencies are represented by arrows
- [ ] Markov chains cannot be extended to deal with continuous time
- [ ] A Markov chain has 2 components: the initial state and the transition matrix
- [ ] A Markov chain is a conditional probability distribution that tells us the probability of the state of the system at time t given the state of the system at time t+1

Maximum marks: 5

14 Markov Models II

Predicting the current state of a system given an initial state and a series of observations up until present is called:

Select one or more alternatives:

- [ ] None of the other options
- [ ] State estimation
- [ ] State prediction
- [ ] Most probable path estimation
- [ ] Smoothing

Maximum marks: 5

15 Hidden Markov Models

Which of the following statements are true?

Select one or more alternatives:

- [ ] The Viterbi algorithm is used to calculate the most probable sequence of state values given an initial distribution and a sequence of observations
- [ ] The Viterbi algorithm is mostly used for statistical analysis of the HMM
- [ ] The long-run distribution is associated to Markov Chains, while the steady-state distribution is associated to a HMM
- [ ] Smoothing the probabilities of a HMM is rarely useful
- [ ] The emission matrix gives the probabilities of observing different emissions given the state of the system

Maximum marks: 5

16 Bayesian Networks I

Which of the following statements are true about Bayesian networks?

- [ ] Any joint probability distribution can be represented as a Bayesian network
- [ ] Conditional independencies cannot be encoded in a Bayesian network
- [ ] The Bayesian Network representation of a joint distribution is unique
- [ ] Bayesian networks are ideal for looking at an event and predicting the likelihood that any of the possible known causes was the contributing factor
- [ ] A joint distribution can be factorised in different manners.

Maximum marks: 5

17 Bayesian Networks II

Which of the following statements are true about Markov Chain Monte Carlo (MCMC) sampling?
Select one or more alternatives:

- [ ] MCMC aims to estimate the probability distributions of a set of unknown variables given a set of known variables
- [ ] A burn period improves the performance of the algorithm and you should always use it
- [ ] MCMC can be used to find the prior distribution and to sample from it
- [ ] Metropolis/Metropolis Hastings and Gibbs Sampler differ on the way they use the samples to generate the desired distributions
- [ ] It is important to always thin your samples

Maximum marks: 5

18 Hyper-parameters

In the context of convolutional neural networks, what is a hyper-parameter?
Select one or more alternatives:

- [ ] Parameters learned from the data
- [ ] Weights and biases are examples of hyper-parameters.
- [ ] Parameters specified by the user or using heuristics.
- [ ] Parameters derived via tranning
- [ ] Final configuration values that are internal to the model

Maximum marks: 5

19 CNNs

Which of the following statements are true?
Select one or more alternatives:

- [ ] The output of a convolutional layer is always smaller than the input if no padding is applied
- [ ] Dropout helps against overfitting by stopping the training when the performance on the validation set begins to deteriorate
- [ ] Convolutional neural networks are particularly useful in applications dealing with image data
- [ ] The input of a convolutional layer is sometimes called a feature map
- [ ] Convolutional layers may be stacked as hidden layers.

Maximum marks: 5

20 Deep Learning

Which of the following statements are true?
Select one or more alternatives:

- [ ] Activation functions are necessary for neural networks to capture nonlinear behaviour
- [ ] Regularisation, increasing the model complexity, and early stopping are all examples of techniques used to solve overfitting
- [ ] Nonlinear functions cannot be obtained by the combination of many linear neurons
- [ ] In machine learning, the test dataset is used to validate the model and adjust the values of hyper-parameters.
- [ ] When a model overfits the data, the mean error on the training dataset increases.

Maximum marks: 5

21 Learning

Which of the following statements are true?
Select one or more alternatives:

- [ ] Convolving 4 different filters of dimension 3x5x5 with an input image of size 3x32x32 without padding produces an output of dimension 4x28x28
- [ ] The loss function computes the distance between the current output and the expected output.
- [ ] In a max pooling layer, the number of learnable parameters depends on the size of the kernel
- [ ] Neural networks are not prone to overfitting
- [ ] A convolutional neural network is trained by maximising the loss function

Maximum marks: 5

22 Hidden Markov Models: Forward-Backward Algorithm

Tables 1 to 4 provide the transition matrix, emission matrix, initial state and a sequence of observations for a hidden Markov model. Use the forward-backward algorithm to calculate the probability distributions for the state of the system at times 0, 1 and 2 given the observations. Show forward values, backward values, and state probabilities.

Table 1: Transition Matrix

S_{t-1}	S_t=0	S_t=1
0	0.7	0.3
1	0.6	0.4

Table 2: Emission Matrix

S	$E=0$	$E=1$
0	0.6	0.4
1	0.9	0.1

Table 3: Initial State

$S=0$	$S=1$
0.5	0.5

Table 4: Observations

Table 1: Observations

Time=1	Time=2
FALSE	FALSE

Fill in your answer here

Maximum marks: 15

23 ReLU

Calculate the output of the convolutional layer using ReLU as activation function

Image

4	10	2
2	0	5
20	1	8

Filter

-1	2
-0.5	2

Fill in your answer here

Maximum marks: 10

24 Text questions

What is the purpose of including randomness in the action-deciding process of a reinforcement learning system? (5 points)

Fill in your answer here

Assume you have a GAN where the discriminator network is a simple binary (Genuine/Fake) classifier. Briefly explain how the generator network is trained. (5 points)

Fill in your answer here

Maximum marks: 10

25 Text questions

Define and compare unsupervised, semi-supervised and one-shot deep learning. (5 points each)

Fill in your answer here

Maximum marks: 15