Exam 2 Weiqi Ji U65196797

(Q1)

a) State search strategies looks at the entire scope of the problem and it uses a function or method to traverse through the states and to find to optimal solution from the start to the goal state. Local search strategies are more focused on a smaller component of the state, and it can produce a solution for this smaller component, therefore breaking down the entire problem state into more digestable chunks.

b) One approach is to train the model first using only datasets that have the attribute to make the model learn the attribute. Then, use the model and train it over the entire dataset, dropping out the data that is marked with the attribute. The downside to this if the attribute was sensitive is that the model may not be able to learn how to properly recognize the attribute.

c)
- caption the video (what is happening)
- learn the actions and be able to categorize these actions later from another video (i.e. camera feed)
- categorize the video itseldf, being able to replicate it or find another like (similar) video.

d) The f-score measure the accuracy of the predicted against the actual results.

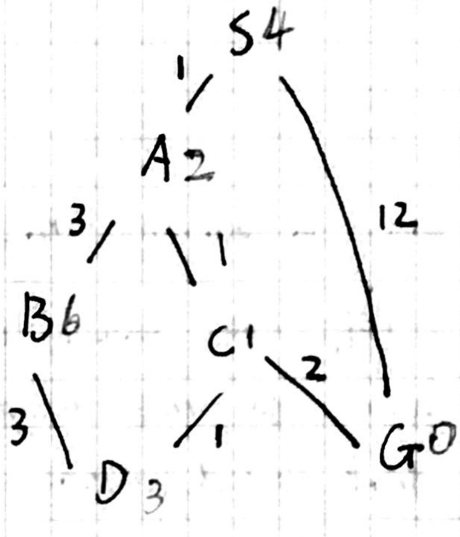
e) i) gradient descent is guarenteed to reach a global maxima if it is computed over multiple mini-batches.

ii) not guarenteed bc it may end up at local maxima

iii). it is made to compute the global optimum, and uses heuristics to help determine it.

f) we ~~use~~ apply a discount to future rewards in reinforcement learning because we want to clarify that the current reward has more weight than future rewards and is more important. Usually it is a factor between 0.9 - 0.99, and it is applied to the sum of all future rewards.

Q2



a) S - G

b) S - A - C - G

c) h_2 is admissible because the distance between the nodes is always $\geq$ than the heuristic.

d) S - G

e) S - A - C - G

(Q3)

a) we want to reduce the spacial res of each feature map during the CNN layers because it saves computing time by having larger features condensed into more compact datasets that still contain the key information of the dataset, which can be passed on to another processing layer.

b) Dropout is very useful against overfitting and it is a regularization technique where random sections of the data is discarded during the computation of the neural network layers.

c) We visualize filters to help us see what features the neural network is highlighting / picking up, and this is useful for use especially in photo and video data.

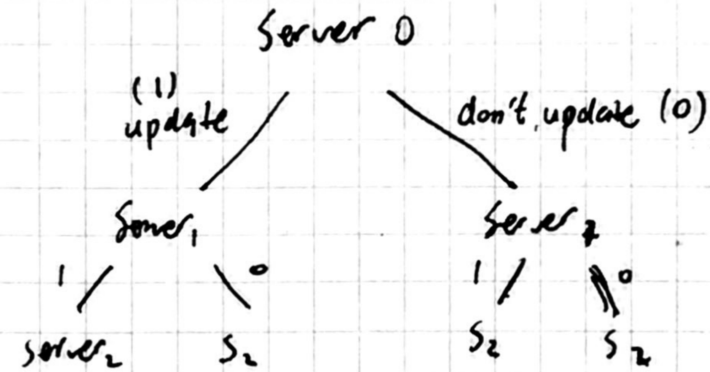
d) i) it is a loss function used for gradient descent

ii)

e) $\Delta \frac{J(\theta)}{x} = \frac{J(\theta)}{\hat{y}} \cdot \frac{\hat{Y}}{J(\theta_2)} \cdot \frac{J(\theta_2)}{\cancel{Z_1}} \cdot \frac{\cancel{Z_1}}{J(\theta_1)} \cdot \frac{J(\theta_1)}{X}$

f) We can use a saliency map during the layers to help see exactly which areas of the video have the most noteable activity and then use that to determine which features we need to model and compute for.

Q4



a) To turn this into a CSP, we must first determine the constraint which is that some servers can be grouped together during the same time instance to be updated or not updated. If we make one of those trees such as above, we can run a DFS on the tree ~~[illegible]~~ to determine possibilities (paths) of satisfactory results to the constraint, if a path violates the constraint, it will not be considered as part of the solution.

b)

T T F

$(\neg a \lor b) \land (\neg b \lor c)$

$\neg a \land b \lor \neg b \land c$

L - N - Q - X

c)

p	q	$p \land q$	$\neg(p \land q)$	$p \lor \neg(p \land q)$	$\neg(\ldots)$
T	T	T	F	T	F
T	F	F	T	T	F
F	T	F	T	T	F
F	F	F	T	T	F

bc the whole expression evaluates to always false, it is a contradiction.

d) $p \to q \longleftrightarrow \neg p \lor q$

$\equiv$

$(\neg p \lor q) \land \neg q \equiv \neg q \land \neg p \lor q \land \neg q$

$\neg(q \lor p) \lor F$

$\equiv$

$\neg(q \lor p) \lor F \to \neg p$

$\equiv$

$\neg(q \lor p) \to \neg p$

$\equiv$

$q \lor \underbrace{p \lor \neg p}_{T}$

Tautology, $p \lor \neg p$ is always T