Q1) B

B contributed to the second AI winter. That was the first question of the assignment.

Q2) C

It was an argument against the Turing Test that it honors the intelligence with regard to humans (Page 24 of the slides).

Q3-Q8

Reference: https://inst.eecs.berkeley.edu/~cs188/fa19/assets/exams/cs188-fa18-mt1.pdf

Q3) A

DFS goes as deep as possible and starts at the left of the tree. Therefore, it explores A, then AA, then AAA, then AAAA, … The first solution that DFS visits will be the one with the most A's on the left, i.e., alphabetically the first match. (AAACCC)

Q4) E

BFS finds the shorted path (at the lowest depth), so the answer should be the shortest password. (CBAC)

Q5) C

Uniform cost search returns the goal with the least cost path. Therefore, it will be the password with the minimum cost. Among the listed password, we have

the cost of option A= AAACCC will be 3X1 +3X3 = 12

the cost option B= ABBCC will be 1+ 2X2 +2X3 = 11

…

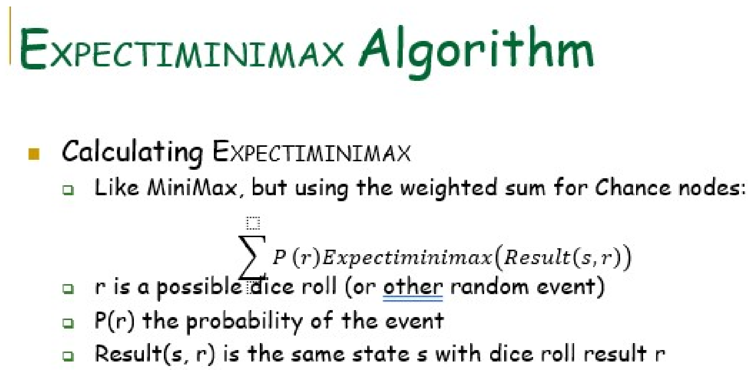
The least cost password will be BABAB = 2+1+2+1+2 = 8 (short and no C since C's are more expensive)

Q6) C

The clue here is that the correct password is unknown, so we cannot have a heuristic that calculates how close we are to the password (to the goal). When there is no heuristic, algorithm A (A*) will not be superior to other algorithms especially when there is no path cost either.

Q7) B

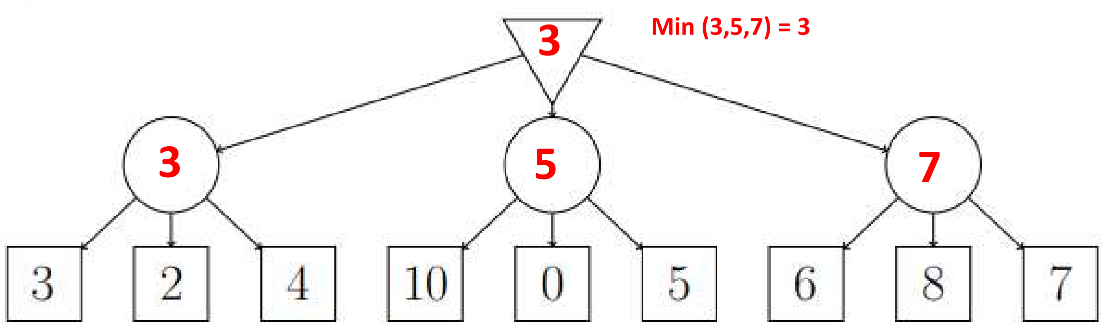
The chance nodes do Expectation meaning the sum of (probability * heuristic values).



The probabilities are all the same, so for the first chance node, the expected value will be (1/3 * 3 + 1/3*2 + 1/3*4) = 3. Note that since probabilities are all the same, it is just simply the average of the three values, i.e., (3+2+4)/3.

For the two other chance nodes, the expected values will similarly be (10+0+5)/3 = 5, and (6+8+7)/3 = 7.

The Min node stills plays Min, so Min(3,5,7) = 3.



Q8) D

Here, we do not know the leaf values in advance, but we know that they are non-negative, and the probabilities are all the same. We start visiting them.

Among the first three nodes, we can not prune anything, since we need all three to calculate the expected value to be 3 (as we saw in Q7).

Now, knowing this 3, our root which is a Min node will secure a 3 from its left branch. It updates its upper bound and from now will accept only values less than 3.

As soon as we explore and see 10 in the second branch, we know that the value that this chance node will send to the root will be >=10/3 (even if we have two zeros, the average of 10 and two zeros will be at least 10/3. If they are not zero, the average increases). Therefore, both leaves (the ones with current values of 0 and 5) can be pruned.

In the last branch, when we explore 6, the average at the chance node will be >= 6/3, so we can not prune at this stage. When we explore 8, the average of this chance node becomes at least (6+8)/3= 14/3 > 3. So regardless of the last leaf, this chance node will not be the chosen branch. The node with the current value of 7 can be pruned.

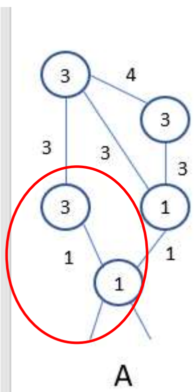
Q9) A

The max(.) heuristic will also be consistent, and consistency is stronger than admissibility, so it will be both consistent and admissible.

Q10) A

To check the consistency, we should check the triangle inequality h(n) <= c(n,n') + h(n-1)

Or we can rewrite it as h(n) – h(n+1) < = c(n, n') (Question 6 of the assignment). In the first graph, we have h(n) = 3, h(n')=1, and C(n,n') = 1 which does not satisfy the triangle inequality.



Q11) D

We discussed the notion of rationality several times in the first two chapters. Rational player in this context means she wants to maximize her own utility, so she wants to minimize the other player's utility (zero-sum game).

Now, if our opponent is not rational, she may not try to minimize our utility, so playing Min-Max would be very conservative and the worst case. So, minimax search is not optimal anymore, and it would be optimal to incorporate some randomness and maybe playing ExpectMiniMax.

Q12) C

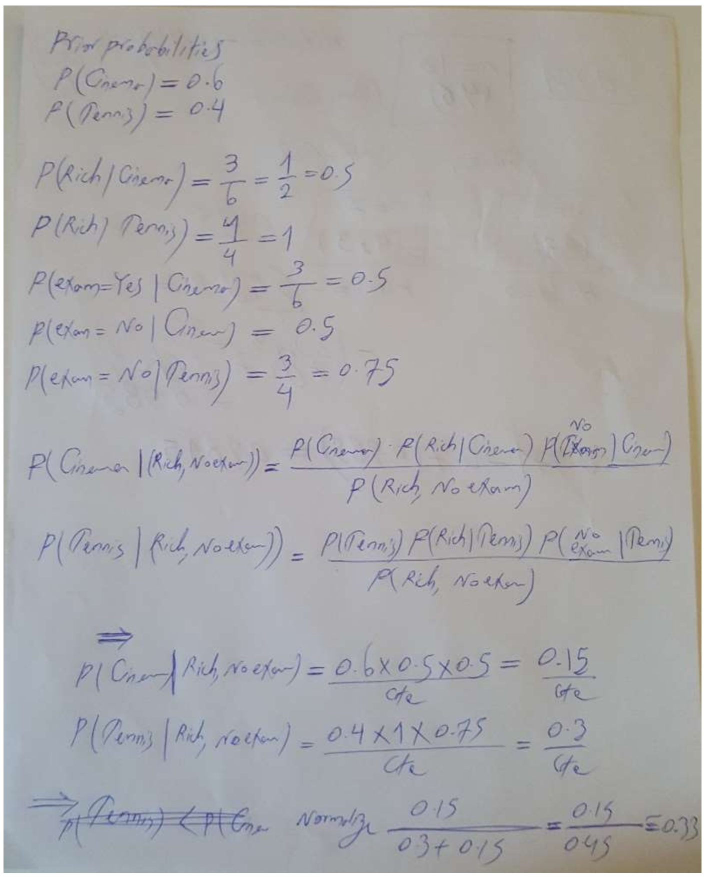
That was the only and the most important assumption in Naïve Bayes.

Q13) D

When we have sparse occurrence meaning when we may observe some events rarely, so without smoothing, their likelihood is considered to be zero.

Q14) E

Solution attached.


Prior probabilities
$P(\text{Cinema}) = 0.6$
$P(\text{Tennis}) = 0.4$

$P(\text{Rich} \mid \text{Cinema}) = \frac{3}{6} = \frac{1}{2} = 0.5$
$P(\text{Rich} \mid \text{Tennis}) = \frac{4}{4} = 1$
$P(\text{exam=Yes} \mid \text{Cinema}) = \frac{3}{6} = 0.5$
$P(\text{exam = No} \mid \text{Cinema}) = 0.5$
$P(\text{exam = No} \mid \text{Tennis}) = \frac{3}{4} = 0.75$

$$P(\text{Cinema} \mid (\text{Rich, No exam})) = \frac{P(\text{Cinema}) \cdot P(\text{Rich} \mid \text{Cinema})\, P(\text{No Exam} \mid \text{Cinema})}{P(\text{Rich, No exam})}$$

$$P(\text{Tennis} \mid (\text{Rich, No exam})) = \frac{P(\text{Tennis})\, P(\text{Rich} \mid \text{Tennis})\, P(\text{No exam} \mid \text{Tennis})}{P(\text{Rich, No exam})}$$

$$\Rightarrow P(\text{Cinema} \mid \text{Rich, No exam}) = \frac{0.6 \times 0.5 \times 0.5}{Cte} = \frac{0.15}{Cte}$$

$$P(\text{Tennis} \mid \text{Rich, No exam}) = \frac{0.4 \times 1 \times 0.75}{Cte} = \frac{0.3}{Cte}$$

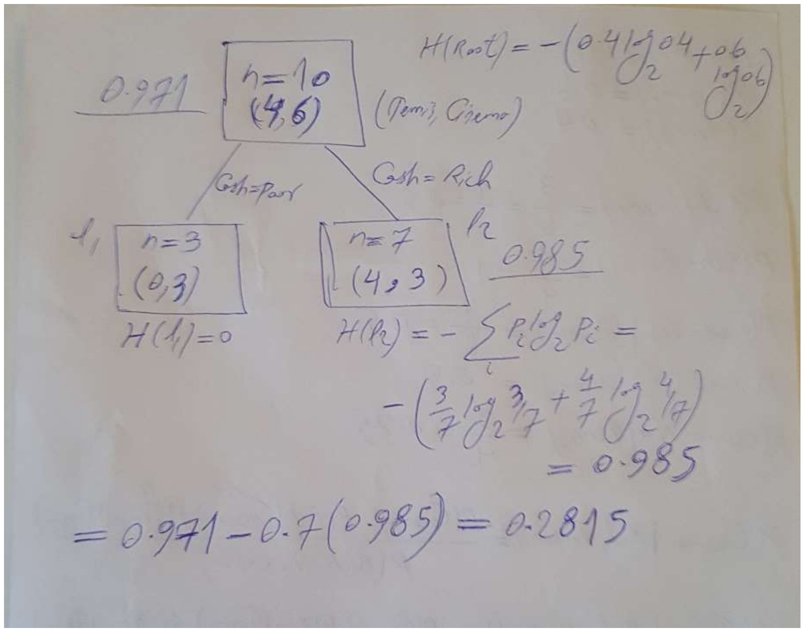
$\Rightarrow \cancel{P(\text{Tennis}) < P(\text{Cinema}}$ Normalize $\frac{0.15}{0.3 + 0.15} = \frac{0.15}{0.45} \cong 0.33$

Q15) B

We thought she either goes to cinema or plays tennis and we predicted a probability. Now, we found that there is also a non-zero probability for shopping, therefore, the share of Cinema decreases (a little of Cinema's share goes to Shopping). (No need to even do any calculation, but we also remember that the probability of (hypothesis | evidence) is found from multiplying the prior probability by the probability of (Evidence | Hypothesis). So, if the prior decreases, the predicted probability will also decrease.

Q16) A

Solution attached.



Q17) D

There are 3 mistakes out of 14, so the accuracy of the tree on training data will be 11/14.

Q18) C

Lower min sample per leaf means we allow a leaf to have fewer samples, so the depth of the tree will increase. The chance of overfitting thus increases.

Q19) Question was eliminated. (1+ bonus point)

Q20) B

Precision is equal to 20/80 = 0.25, and recall is 20/100 = 0.2. F1 score is thus 0.22

Q21) A

It may improve the bias but may increase the chance of overfitting.

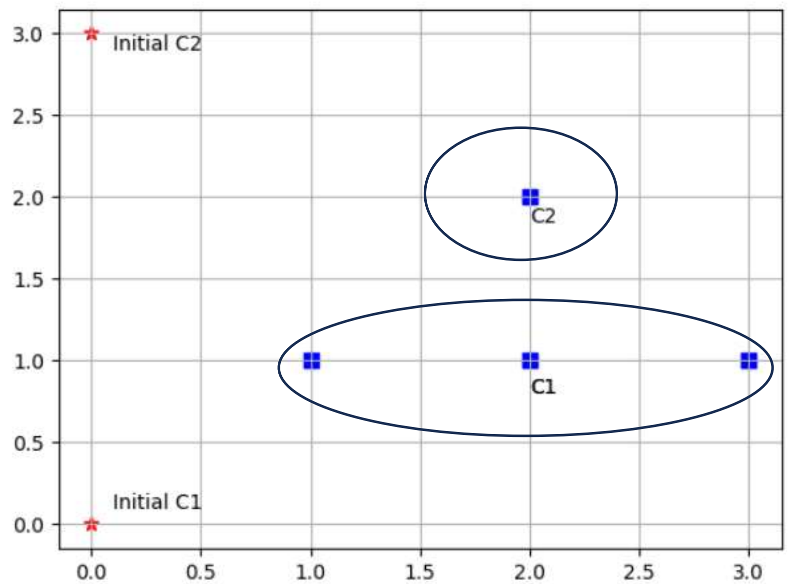
Q22) E

(1,1), (2,1), (3,1) are all closer to (0,0) than to (0,3)

They will join the cluster with centroid of (0,0).

(2,2) is closer to (0,3).

The new centroid will be (2,1) for the cluster with 3 observations. For the cluster with one observation, itself will be the centroid, so (2,2). All observations are stable and do not move.



Q23) B

In Agglomerative Clustering, we do not start with centroids. We start with each sample as one cluster.

Q24) D

Model 3 has always a higher complexity than model 1.

In the average approach, we find the distance between every pair of nodes while in ‘single’ and ‘complete’ methods, we find the distance between the two nearest and farthest nodes only (Finding nearest and farthest are also costly but with a good implementation, we can do it efficiently).

Q25) E

It depends on the performance of the classifier. We do not know how many predictions will have more than 90% or less than 10% confidence.