COMP 4190 Assignment 2 Answer

Fidelio Ciandy, 7934456

Problem 1

For my first question I have four different functions

- **queens_puzzle**
  - This function initializes the board with empty cells
  - It initializes a set to store the queens positions
  - It initializes an array to store the solutions
  - It calls the backtrack function
- **backtrack**
  - This function is where the backtracking happens
  - For each row, it recursively iterates through each column and checks whether placing a queen at that cell is valid
  - If the queen can be attacked, it continues to the next column
  - If we successfully reach the last row, we add that board to the solution
- **is_attacked**
  - This function will iterate through the queens set and see if the current queen can be attacked by any queen in that set or not
- **determine-queen-attack**
  - This function will check whether two different queen position can attack each other

Problem 2

- **words-from-board**
  - This function will iterate through each word, and the board
  - Check whether if the current word matches any word on the board
  - Will use the backtrack algorithm to check whether the word match or no
- **backtrack**
  - This function is where the backtracking happens
  - Will iterate through each cell neighbor

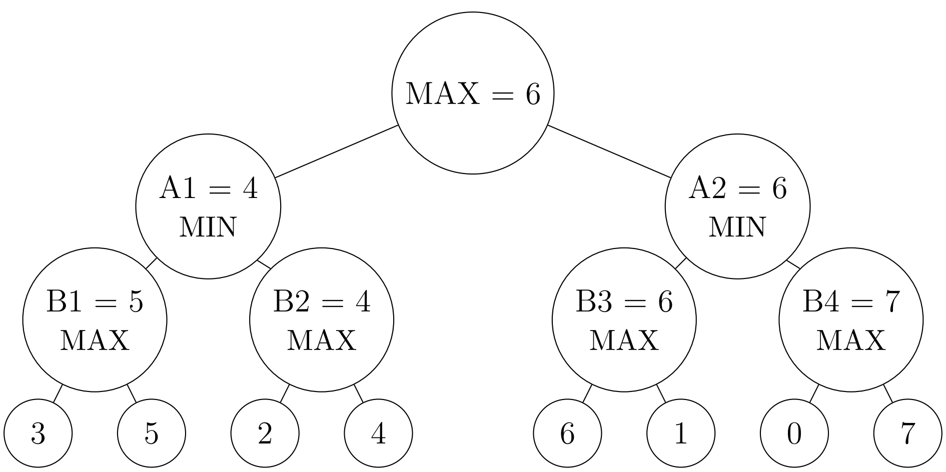
  - Will check if the cell has been visited or not
  - For each neighbor, if the character is still the same as the word,
  - It will keep going deeper,
  - else, it will backtrack

- **FindNeighbors**
  - Given a specific entry (row, col), this function will calculate the neighbors for that specific (row, col)

Problem 3

- **piles**
  - Using dynamic programming,
  - first, it will initialize the calculation when there's only 1 pile left
  - calculating and saving for each possible pile
  - Then, it will iterate through each possible pile, e.g., 2,3,4,..n.
  - And calculate the outcome of when we take the left pile or the right pile
  - will choose the best possible outcome (max)
  - Lastly, if we have calculated each entry, if the last entry is bigger than 0.
  - That means that Alice wins the game

Problem 4

(a) 

(b) - Based on my computation, MAX should choose the right side at the root.
- 6 is the payoff MAX can guarantee despite MIN's action
- minimax corresponds to a worst-case guarantee for MAX because just below MAX, there's MIN which forces MAX to choose the best out of the worst (MAX's worst case). Thus, a worst-case guarantee.

(c) Optimization

(i) The reason why performing a min after max gives min more control over the final outcome is because if minimization is done after maximization, and the min moves second, thus, will choose the worst possible outcome for the last maximizer. By choosing the worst possible outcome for the max, minimizer can force the final value as small as possible. Thus, gives the minimizing player more control over the final outcome.

(ii)

$$\begin{aligned} V &= \max_a \min_b \max_{c \in children(b)} LeafValue(a,b,c) \\ &= \max_a \min_b g(a,b) \end{aligned}$$

Lets try to compute the tree using the formula

$$V = \max_{A1,A2}(\min_{B1,B2}(g(A1,B1), g(A1,B2)), \min_{B3,B4}(g(A2,B3), g(A2,B4)))$$

$$= max(min(max(3,5), max(2,4)), min(max(6,1), max(0,7)))$$

$$= 6$$

(iii)

compare $\max_a \min_b g(a,b)$ and $\min_b \max_a g(a,b)$

Compute $\rightarrow \min_b \max_a g(a,b)$

Since min goes last but the "max" doesn't know what value min is going to choose, there could be 4 different choices that min will choose:

i. B1, B3
ii. B1, B4
iii. B2, B3
iv. B2, B4

Thus we define:

$$b \in \{(B1,B3),(B1,B4),(B2,B3),(B2,B4)\}, \quad a \in \{A1, A2\}$$

For each such pair $b = (b_1, b_2)$, we compute:

$$\max_a \ (g(A1,b_1),\ g(A2,b_2))$$

i. **Evaluate:** $b = (B1, B3)$

$$\max_a(g(A1,B1),\ g(A2,B3)) = max(max(3,5), max(6,1)) = 6$$

ii. **Evaluate:** $b = (B1, B4)$

$$\max_a(g(A1,B1),\ g(A2,B4)) = max(max(3,5), max(0,7)) = 7$$

iii. **Evaluate:** $b = (B2, B3)$

$$\max_a(g(A1,B2),\ g(A2,B3)) = max(max(2,4), max(6,1)) = 6$$

iv. **Evaluate:** $b = (B2, B4)$

$$\max_a(g(A1, B2),\ g(A2, B4)) = max(max(2, 4), max(0, 7)) = 7$$

After calculating all this, min will choose the minimum value, thus $Vmin = 6$
From part (2), the game value V is also 6. Thus, they are equal

(iv) Suppose two quantities above are equal. This is what's called the Minimax Theorem, the one that's being shown in the slides. It implies that revealing the optimal strategy doesn't hurt you. Meaning that, neither player has any benefits from changing strategy or moving order. Thus, the game has a stable optimal strategies.

Problem 5

(a)

$$\sum_X \sum_Y P(X = x, Y = y) = 0.10 + 0.20 + 0.10 + 0.15 + 0.25 + 0.20 = 1$$

(b)

$$P(X = 0) = 0.10 + 0.20 + 0.10 = 0.40$$

$$P(X = 1) = 0.15 + 0.25 + 0.20 = 0.60$$

$$P(Y = 0) = 0.10 + 0.15 = 0.25$$

$$P(Y = 1) = 0.20 + 0.25 = 0.45$$

$$P(Y = 2) = 0.10 + 0.20 = 0.30$$

(c) Conditional probability formula:

$$P(X \mid Y) = \frac{P(X \cap Y)}{P(Y)}$$

Bayes' theorem:

$$P(X \mid Y) = \frac{P(X)\, P(Y \mid X)}{P(Y)}$$

$$P(X = 1 \mid Y = 1) = \frac{0.6(0.25/0.6)}{0.45} = 0.56$$

$$P(Y = 2 \mid X = 1) = \frac{0.3(0.20/0.3)}{0.6} = 0.33$$

$$P(X = 0 \mid Y = 0) = \frac{0.10}{0.25} = 0.40$$

(d)

$$P(X = 0 \mid Y = 1) = \frac{0.2}{0.45} = 0.44$$

$$P(X = 1 \mid Y = 1) = \frac{0.25}{0.45} = 0.56$$

$$P(Y = 0 \mid X = 0) = \frac{0.1}{0.4} = 0.25$$

$$P(Y = 1 \mid X = 0) = \frac{0.2}{0.4} = 0.5$$

$$P(Y = 2 \mid X = 0) = \frac{0.1}{0.4} = 0.25$$

(e) Conditional probability identity:
if $P(A|B) = P(A)$ and $P(B|A) = P(B)$, then events A and B are independent.
Seeing from $P(X = 0 \mid Y = 1) = \frac{0.2}{0.45} = 0.44$ and $P(X = 0) = 0.10 + 0.20 + 0.10 = 0.40$
$0.44 \neq 0.40$, thus, X and Y are not independent.

(f) 1. Joint probability $P(X, Y) \rightarrow$ the probability that both happens at the same time. In other words, the probability that the AI model correctly predicts "object" at a specific sensor. For instance, $P(X = 1, Y = 1)$, meaning that probability of AI model predict "yes" and the object actually present at Y = 1 is 0.2

2. Marginal probability $P_x(X) \rightarrow$ the probability of X alone, ignoring Y. In other words, the probability that the object present or not present, independent of Y.

3. Conditional probability $P(X|Y) \rightarrow$ the probability of X given that Y already happened. In other words, the probability of the AI model predicts "object present" or not given an object has been detected previously at Y. For instance, $P(X = 1|Y = 1)$, is the probability that AI predicts an "object present" given object detected by sensor 1.

Problem 6

(a) Suppose generate 1000 samples

(1) Expected samples from each pair (X,Y) out of 1000 samples

(I) $(X = 0, Y = 0) = 1000 * 0.1 = 100$
(II) $(X = 0, Y = 1) = 1000 * 0.2 = 200$
(III) $(X = 0, Y = 2) = 1000 * 0.1 = 100$
(IV) $(X = 1, Y = 0) = 1000 * 0.15 = 150$
(V) $(X = 1, Y = 1) = 1000 * 0.25 = 250$
(VI) $(X = 1, Y = 2) = 1000 * 0.2 = 200$

(2) Samples where X = 0. Y = 1

(I) $X = 0 \rightarrow 1000 * (0.1 + 0.2 + 0.1) = 400$
(II) $Y = 1 \rightarrow 1000 * (0.2 + 0.25) = 450$

(b) Simulate sampling,

(I) by placing 1000 cards in a box. Each card has label (X, Y), where the number of each card match the expected samples from (a)

(II) mix those 1000 cards

(III) draw one card at random from the box. This represents one sample from this box

(c) The reason why this simulates to conditional probability $P(X|Y)$ is because, by getting the value of Y and restrict the sample only by using that same value of Y. Meaning that we randomly selects only from the reduced set of samples using that same value of Y. Thus, this relates to the definition of conditional probability, where the probability of getting X from this reduced set represents $P(X|Y)$, which is the probability of X, given that Y has occured.

(d) Suppose that we generate more and more sample from this distribution. Note the formula of standard deviation. Low standard deviation means that data is consistent. On the other hand, high standard deviation means data has high variability. Since standard deviation is inversely propotional from N, thus, as sample N increases, the standard deviation decreases. Thus, data is more consistent, and should approach the probabilities in the table.

Problem 7

(a)

$$\int_{-\infty}^{\infty} C\ exp(-\frac{(x-\mu)^2}{2\sigma^2})dx = 1$$

Let $u = x - \mu$

$$\frac{du}{dx} = \frac{d}{dx}(x-\mu)$$

$$\frac{du}{dx} = 1$$

$$du = dx$$

$$\int_{-\infty}^{\infty} C\ exp(-\frac{u^2}{2\sigma^2})du = 1$$

$$C\int_{-\infty}^{\infty} exp(-\frac{u^2}{2\sigma^2})du = 1$$

Let a $= \dfrac{1}{2\sigma^2}$

Solve for $\rightarrow \displaystyle\int_{-\infty}^{\infty} exp(-au^2)du$

Using the Gaussian Integral Property, the above calculation will simplify to

$$\int_{-\infty}^{\infty} exp(-au^2)du = \sqrt{\frac{\pi}{a}}$$

Prove of Gaussian Integral Property:

Let $I = \displaystyle\int_{-\infty}^{\infty} exp(-ax^2)dx$

Integrate over both x and y:

$$I^2 = \int_{-\infty}^{\infty} exp(-ax^2)dx \int_{-\infty}^{\infty} exp(-ay^2)dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} exp(-a(x^2+y^2))dxdy$$

Transform to polar coordinates:

$$= 2\pi \int_{0}^{\infty} rexp(-ar^2)dr$$

$= \pi \int_0^{\infty} d(r^2) exp(-ar^2)$

$= \pi \left[-\frac{1}{a} e^{-ar^2} \right]_0^{\infty} = \frac{\pi}{a}$

Take the root square

$= \sqrt{\frac{\pi}{a}}$

Now, solve for C

$C \sqrt{\frac{\pi}{a}} = 1$

$C \sqrt{\frac{\pi}{\frac{1}{2\sigma^2}}} = 1$

$C = \frac{1}{\sigma\sqrt{2\pi}}$

(b)

$E[X] = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \ exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$

$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \ exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$

Let $u = x - \mu$

$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (u+\mu) \ exp(-\frac{u^2}{2\sigma^2})dx$

$= \frac{1}{\sigma\sqrt{2\pi}} [\int_{-\infty}^{\infty} u \ exp(-\frac{u^2}{2\sigma^2})du + \mu \int_{-\infty}^{\infty} \ exp(-\frac{u^2}{2\sigma^2})du]$

Since, $u \ exp(-\frac{u^2}{2\sigma^2})$ is an odd function. By even and odd integral theorem,

$\int_{-\infty}^{\infty} u \ exp(-\frac{u^2}{2\sigma^2})du = 0$

Thus,

$= \frac{1}{\sigma\sqrt{2\pi}} [0 + \mu \int_{-\infty}^{\infty} \ exp(-\frac{u^2}{2\sigma^2})du]$

From the question (a), we can simplify this to

$= \frac{1}{\sigma\sqrt{2\pi}} [\mu\sigma\sqrt{2\pi}]$

$= \mu$

(c)

$Var(X) = E[(X-\mu)^2]$

$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \ exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$ ..From part (b)

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 \; exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx$$

Let $u = x - \mu$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 \; exp(-\frac{u^2}{2\sigma^2}) du$$

Let a $= \frac{1}{2\sigma^2}$ Solve for:

$$\int_{-\infty}^{\infty} u^2 \; exp(-au^2) du$$

Using Integral of Gaussian Properties

$$\int_{-\infty}^{\infty} u^2 \; exp(-au^2) du = \frac{1}{2a}\sqrt{\frac{\pi}{a}}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{2a}\sqrt{\frac{\pi}{a}}$$

Subtitute back a:

$$= \frac{1}{\sigma\sqrt{2\pi}} \sigma^3\sqrt{2\pi}$$

$$= \sigma^2$$

(d)

$$Z = \frac{X-\mu}{\sigma}$$

$$E[Z] = E[\frac{X-\mu}{\sigma}]$$

$$= \frac{1}{\sigma} E[X-\mu]$$

$$= \frac{1}{\sigma} (E[X]-\mu)$$

$$= \frac{1}{\sigma} (\mu-\mu) \text{ ... From previous question (b)}$$

$$E[Z] = \mu = 0 \;\; ...(1)$$

$$Var(Z) = E[(Z-\mu)^2]$$

$$Var(Z) = E[(Z-0)^2] \text{ ... From (1)}$$

$$Var(Z) = E[(Z)^2]$$

$$Var(Z) = E[(\frac{X-\mu}{\sigma})^2]$$

$$Var(Z) = \frac{1}{\sigma^2} E[(X-\mu)^2]$$

$$Var(Z) = \frac{1}{\sigma^2}\sigma^2 \text{ ... From part (c)}$$

$$Var(Z) = 1 \;\; ...(2)$$

Thus, from (1) and (2), Z follows the standard normal distribution with mean 0 and variance 1

(e) (I) $\mu \rightarrow$ represents the average of whether is the noise represent the true value or no. In other words, the average error of predicting the noise.

(II) $\sigma^2 \rightarrow$ represents the degree of spread, whether the prediction is certain or not.
If high $\rightarrow$ meaning lot of spread in the prediction, not certain.
Else, if low $\rightarrow$ meaning less spread in the prediction, more certain with the noise prediction.

(III) Why larger σ^2 means higher uncertainty? As mention previously, higher variance means that data are widely spread out, predictions of noises are vary. Thus, it's more uncertain in predictions.