IN4050/3050 Sample Questions for Final Exam

Pooya Zakeri

- Consider the problem of predicting whether a loan application will be approved (Yes/No) based on applicant information using the following dataset:

ApplicantID	Gender	Marital Status	Education	Employment Type	Income	Loan Amount	Property Area	Loan Status
1001	Male	Married	Graduate	Salaried	High	Large	Urban	Yes
1002	Female	Single	Non-Grad	Self-Employed	Low	Medium	Rural	No
1003	Male	Married	Graduate	Salaried	High	Medium	Semi-Urban	Yes
1004	Female	Married	Non-Grad	Salaried	Medium	Small	Urban	No

What is the most common algorithm to solve this problem given the types of features?

A. kNN
B. Perceptron
C. Decision trees
D. Deep Neural Network such as ResNet-150

Solution: C

- Decision Trees
- This problem is:
  - Binary classification (Yes/No)
  - Based on tabular data (applicant information such as income, age, history, etc.)
  - Features are categorical
- For tabular data, the most commonly used and best-performing classical algorithm is Decision Trees
  - (and their ensembles like Random Forest)

- Suppose we use Partially Mapped Crossover (PMX) on the following two parent tours with a randomly chosen cut point between the 3rd and 4th bits and another cut point between the 6th and 7th bits:
  - Parent 1: P1 = (3, 4, 8, |2, 7, 1,| 6, 5)
  - Parent 2: P2 = (4, 2, 5, |1, 6, 8,| 3, 7)

  - *Hint: Partially Mapped Crossover (PMX) algorithm:*
    - *1. Substring selection: Cut each parent into two substrings, and then select one substring for each parent at random.*
    - *2. Substring exchange: Exchange the two selected substrings to produce proto-offspring.*
    - *3. Mapping list determination: Determine the mapping relationship based on the selected substrings.*
    - *4. Offspring legalization: Legalize proto-offspring with the mapping relationship.*

What offspring can PMX generate from these parents? (Choose TWO answers)

A. A) (3, 4, 5, 2, 7, 1, 6, 8)
B. B) (4, 8, 5, 2, 7, 1, 3, 6)
C. C) (4, 2, 5, 1, 6, 8, 3, 7)
D. D) (4, 2, 5, 2, 7, 1, 3, 5)
E. E) (4, 2, 5, 1, 6, 8, 3, 7)
F. F) (3, 4, 2, 1, 6, 8, 7, 5)
G. G) (4, 2, 7, 1, 6, 8, 5, 3)
H. H) (3, 4, 8, 2, 7, 1, 6, 5)

Solution: B and F

- For which of the following problems is a genetic algorithm generally *not* very efficient?

  - Hint:
  - *N-Queens Problem: For an NxN chess board, place N queens such that no queen is able to attack another.*
  - *Job-Shop Scheduling : Assigning a series of jobs to resources (like machines) while minimizing total completion time and ensuring no overlap for jobs requiring the same machine.*
  - *Graph Coloring Problem: Coloring the graph vertices in such a way that no two adjacent vertices have the same color.*
  - *Knapsack Problem: Given a set of items, each with a weight and a value, determine which items to include in the collection so that the total weight is less than or equal to a given limit and the total value is as large as possible.*

- A) Finding Large Prime Numbers
- B) N-Queens Problem (for large N)
- C) Job-Shop Scheduling
- D) Graph Coloring Problem
- E) Knapsack Problem

Solution: A) Finding Large Prime Numbers

- Genetic Algorithms are generally effective in discrete optimization problems, where the solution space is made up of distinct, separate options. However, when the fitness landscape (the way the quality of solutions is evaluated) is not continuous, GAs struggle. For example, using a GA to find a large prime number wouldn't work well because the fitness function in that case is discontinuous; it jumps from non-prime to prime without gradual improvement.

- In a genetic algorithm, a population has the following fitness values for five individuals: [12, 15, 20, 25, 28]. Calculate the probability of selecting the individual with a fitness of 28 using fitness-proportionate selection (roulette wheel).

- A) 0.14
- B) 0.28
- C) 0.4
- D) 0.56

Solution: B

- Fitness values: $12,15,20,25,28$

Compute total fitness

  - $12 + 15 + 20 + 25 + 28 = 100$
- Probability of selecting the individual with fitness 28
  - $P = \frac{28}{100} = 0.28$

- You are using a roulette wheel selection method in a genetic algorithm for a problem with a fitness landscape where most individuals have similar fitness values. What issue might arise with this selection method, and which alternative selection strategy would address this issue?

- A) Premature convergence; switch to tournament selection
- B) Loss of selection pressure; switch to rank-based selection
- C) Increased diversity; switch to elitist selection
- D) Lack of convergence; switch to random selection

Solution: B

- If everyone has almost the same fitness, the probabilities of being selected are nearly equal (Loss of selection pressure). This means good individuals are not favored strongly enough. Evolution becomes slow or even stalls.
- Rank-based selection assigns selection probabilities based on the *rank* of the individual, not the raw fitness values. So even if fitness values are very close, the best individual still gets a much higher selection probability. The worst individual gets a much lower probability.

- Which of the following statements about tournament selection in a genetic algorithm are true? (Select TWO)

  - A) Increasing tournament size reduces selection pressure.
  - B) Larger tournament sizes favor fitter individuals.
  - C) Tournament selection has constant selection pressure regardless of tournament size.
  - D) Smaller tournament sizes increase diversity.
  - E) Tournament selection is guaranteed to select the top individual.
  - F) Picking with replacement increases selection pressure.

Solution: B and D

- B) Larger tournament sizes favor fitter individuals. True
- In tournament selection, you randomly sample *k* individuals. The larger the tournament size (k), the more likely a very fit individual will be included and win. This increases selection pressure toward fitter individuals.
- D) Smaller tournament sizes increase diversity. True
- With a small tournament size, even weak individuals have a chance to win. This reduces selection pressure and increases genetic diversity in the population.
- A) Increasing tournament size reduces selection pressure. False
  It increases selection pressure.
- C) Tournament selection has constant selection pressure regardless of tournament size. False
  Selection pressure depends directly on tournament size.
- E) Tournament selection is guaranteed to select the top individual. False
  Only guaranteed if tournament size equals population size (which is almost never used).
- F) Picking with replacement increases selection pressure. False (at least not generally true)
- Tournament selection usually samples without replacement.
- With replacement doesn't systematically increase selection pressure; the key factor is tournament size, not sampling method

- For tournament selection with a sample size k=3 in a population of 6 individuals, calculate the probability that the individual with rank 1 is selected in a single tournament (assuming no replacement.)
  - A) 0.2
  - B) 0.3
  - C) 0.33
  - D) 0.5

Solution: D

- Problem Breakdown
- Sample size (k): 3
- Population size: 6
- *Selection without replacement: Once an individual is selected, they cannot be selected again in the same tournament.*
- *Goal: Calculate the probability that the individual with rank 1 (the best individual) is included in the sample of 3 and wins the tournament.*
- Calculate the Total Number of Ways to Select 3 Individuals from 6:
  - Total *ways to select 3 individuals* $= \binom{6}{3} = \frac{6!}{3!(6-3)!} = 20$
- Calculate the Number of Favorable Outcomes:
  - To ensure that the individual with rank 1 is selected as the winner, we must include them in the sample of 3 and ensure that they are the highest-ranked individual in that sample.
  - If the individual with rank 1 is included, we need to select the remaining 2 individuals from the other 5 individuals in the population (ranks 2 to 6).
- The number of ways to include rank 1 and choose 2 additional individuals from the remaining 5 is:
- Favorable *combinations* $= \binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$
- Calculate the Probability:
  - Probability that rank 1 is included in a sample of 3 (and thus the highest-ranked in that sample) is:
- $$P(\textit{rank 1 is selected}) = \frac{\textit{Favorable combinations}}{\textit{Total combinations}} = \frac{10}{20} = 0.5$$

- If an evolutionary algorithm applies fitness sharing with a niche size σ_share=2, and three individuals with fitness values [9, 8, 7] are located within a distance less than σ_share, calculate the shared fitness for the individual with fitness 8. Assume sharing function sh(d)=1−d/σ_share.

  - A) 2.6
  - B) 4
  - C) 6
  - D) 8.1

Solution: B

- **Niche size: $\sigma_{sh} = 2$**
- **Three individuals are all within distance < σ_{sh}**
- **Fitness values: [9, 8, 7]**
- **Sharing function:**
  - $sh(d) = 1 - \frac{d}{\sigma_{share}}$
  - **For the individual itself: $d = 0$**
  - $sh(0) = 1 - 0 = 1$
  - **For the two neighbors at distance $d = 1$:**
  - $sh(1) = 1 - \frac{1}{2} = 0.5$
  - **So the niche count (sum of sharing values for that individual with fitness 8) is:**
  - $m = 1 + 0.5 + 0.5 = 2$
  - **Then the shared fitness is:**
  - $f' = \frac{f}{m} = \frac{8}{2} = 4$

- For a binary classifier, the precision-recall curve is particularly useful compared to the ROC curve when:
  - A) Classes are balanced, providing equal distribution of positives and negatives.
  - B) The model requires a balance between sensitivity and specificity.
  - C) The dataset is highly imbalanced with a rare positive class.
  - D) The classifier has high accuracy but low AUC.

Solution: C

- ROC curves can look deceptively good when negatives vastly outnumber positives.
- Precision–Recall curves focus only on the positive class (precision = TP / (TP + FP)), so they are far more informative when positive examples are rare.

- Which of the following statements are true? (Select TWO)
  - A) Ridge regularization is likely to lead to a sparse model by shrinking some coefficients exactly to zero.
  - B) Lasso regularization is likely to lead to a sparse model by shrinking some coefficients exactly to zero, which is helpful for feature selection.
  - C) Early stopping is likely to lead to a sparse model by reducing model complexity.
  - D) Adding more data to a model with high bias is likely to reduce the model's overfitting.
  - E) Adding more data to a model with high variance can sometimes reduce the model's overfitting.

Solution: B and E

- A) Ridge regularization is likely to lead to a sparse model…
  False. Ridge (L2) shrinks coefficients toward zero but almost never exactly to zero, so it usually does not produce a sparse model.
- B) Lasso regularization is likely to lead to a sparse model…
  True. Lasso (L1) can shrink some coefficients exactly to zero, which effectively performs feature selection and gives a sparse model.
- C) Early stopping is likely to lead to a sparse model…
  False. Early stopping is a regularization method that limits overfitting by stopping training early, but it doesn't typically produce sparsity in the sense of zeros in the parameters.
- D) Adding more data to a model with high bias…
  False. High bias = underfitting. Adding more data usually doesn't fix underfitting; you need a more flexible model or less regularization.
- E) Adding more data to a model with high variance can sometimes reduce the model's overfitting.
  True. High variance = overfitting. More data helps the model generalize better and can reduce overfitting.

- What is the primary difference between Strong AI and Weak AI? (Choose TWO)
  - A) Strong AI claims AI can truly understand and have minds.
  - B) Weak AI is limited to performing specific tasks without understanding.
  - C) Strong AI is less computationally intensive.
  - D) Weak AI seeks to create machines with consciousness.

Solution: A and B



- The "Chinese Room" argument by John Searle challenges which assumption in AI?
  - A) That AI can replicate human emotions
  - B) That AI can only perform narrow, specific tasks
  - C) That symbolic AI is inferior to connectionist approaches
  - D) That computation alone is sufficient for understanding

Solution: D

- How does the complexity of an AI model generally affect its explainability, and why is this relationship important in the context of Explainable AI? (Select TWO)
  - A) As model complexity increases, explainability typically decreases, making it harder for humans to understand the decision-making process.
  - B) Complex models, such as deep neural networks, are usually easier to explain than simpler models like decision trees.
  - C) Explainable AI techniques are more effective on complex models than on simple models.
  - D) Simple models, like linear regression, are usually less transparent, making them more difficult to explain compared to more complex models.
  - E) Lack of explainability in complex AI models can lead to ethical issues, as it becomes difficult to identify and correct biases in decisions that impact

Solution: A and E

- Choose **two correct statements** among all statements:
  - Pure exploitation may find a local optimum by gradually improving solutions
  - until we can’t find a better solution
  - Global optimum can be worse than some neighboring solutions.
  - Exploration means moving (randomly) around in the search landscape, trying to find new
  - types of solutions that potentially are better than what we have found before.
  - Local optimum can be worse than some neighboring solutions.

Solution: A and C

- “A” describes exactly how greedy/exploitative search works: it keeps improving a solution until it reaches a local optimum, where no neighboring solution is better.
- A global optimum is the best solution in the entire search space, so it cannot be worse than any neighbor.
- Exploration is about searching new regions of the solution space, often using randomness, to escape local optima.
- By definition, at a local optimum all neighbors are worse or equal. If a neighbor was better, it wouldn’t be a local optimum.

- You have a neural network with:
- Inputs X
- One hidden layer with activations H
- Output layer with predictions Y
- During training, you compute:

```
# Forward pass:
H = hidden_forward(X)      # shape: (N, H)
Y = output_forward(H)      # shape: (N, 1)

# Deltas:
out_deltas = Y - T         # T = targets, shape: (N, 1)
hidden_deltas = out_deltas @ W2.T * hidden_activation_derivative(H)
```

Which weight update formulas are correct?

Here:
- `W1` are the weights from input → hidden
- `W2` are the weights from hidden → output
- `lr` is the learning rate

A
```
W2 -= lr * H.T @ out_deltas
W1 -= lr * X.T @ hidden_deltas
```

B
```
W2 -= lr * X.T @ out_deltas
W1 -= lr * H.T @ hidden_deltas
```

C
```
W2 -= lr * out_deltas.T @ H
W1 -= lr * hidden_deltas.T @ X
```

D
None of the above

Solution: A

- Because for each layer:
  - weights_l -= lr * (inputs_to_layer_l).T @ (deltas_at_outputs_of_layer_l)
- For W2 (hidden → output):
  inputs = H, deltas = out_deltas → H.T @ out_deltas
- For W1 (input → hidden):
  inputs = X, deltas = hidden_deltas → X.T @ hidden_deltas

- We have a dataset consisting of the numpy arrays X and t, and a multinomial logistic regression classifier with the following forward function:
- def forward(self, X):
  X = self.add_bias(X)
- return softmax(X @ self.weights)
- In addition, we get the following prints:
- $ print(X.shape)
- > (125, 3)
- $ print(cl.weights.shape)
- > (4, 8)
- How many classes do we have?

a) 3
b) 4
c) 8
d) 12

Solution: C

- In multinomial logistic regression, the second dimension of the weight matrix (here, 8) is the number of output units/classes, so the model predicts **8 classes**.

- Which of the following best contrasts *compression* vs *dimensionality reduction* as presented?

- A) Compression is always lossless; dimensionality reduction is always lossy
  B) Both aim at human-interpretable plots
  C) Compression aims at compact encoding with the ability to decode, while dimensionality reduction aims to preserve structure for interpretation/analysis (decoding is not the main goal)
  D) They’re identical in goal

Solution: C

- Compression focuses on a compact representation that can be decoded back to (an exact or approximate) original. Dimensionality reduction focuses on preserving the important structure or relationships in the data (often for visualization or further analysis) without requiring that you can reconstruct the original data perfectly.

Functionalism defines mental states primarily by...

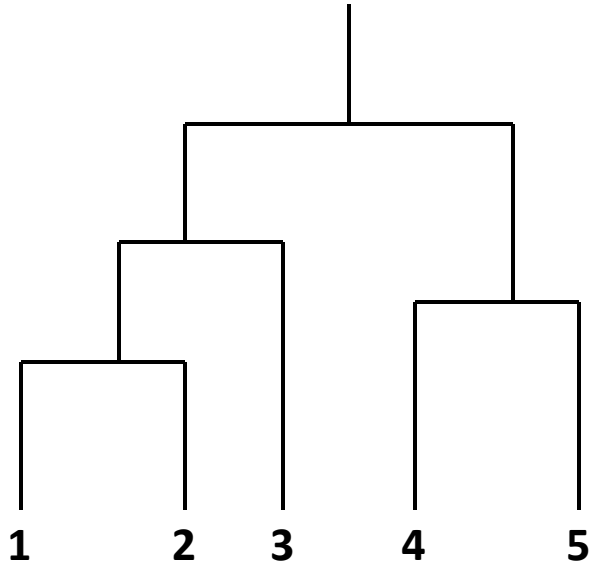
A) Their physical composition
B) Their emotional content
C) Their causal roles and functional organization
D) Their location in the brain

Solution: C

- In functionalism, what makes something a mental state is **how it functions;** its causal role in the system (how it processes inputs, interacts with other states, and produces outputs)—not what it's physically made of or where it's located.

- Consider five items 1,2,3,4,5with the (symmetric) cosine similarity matrix shown below. Using these similarities, you perform agglomerative hierarchical clustering and obtain the dendrogram shown.

	I1	I2	I3	I4	I5
I1	1.00	0.90	0.10	0.65	0.20
I2	0.90	1.00	0.70	0.60	0.50
I3	0.10	0.70	1.00	0.40	0.30
I4	0.65	0.60	0.40	1.00	0.80
I5	0.20	0.50	0.30	0.80	1.00



Which linkage method is most consistent with the dendrogram?

- A) Single linkage
  B) Average linkage
  C) Ward's linkage
  D) Complete linkage
  E) Either single or average linkage are both consistent with this dendrogram

Correct answer: E

- Using cosine similarity, both **single** (max similarity / min distance) and **average** linkage lead to the same merge order and cluster structure as in the dendrogram (e.g., item 3 is more similar to cluster {1,2} than to {4,5} under both definitions). Ward and complete linkage would generally produce different merge heights/ordering for these similarities, so the most correct statement is that both single and average linkage are consistent.