

Amazon ML Summer School India

Previous Year Questions Detailed Solutions Step by Step

Prepared by Vikram Gaur



The selection test will have two parts – Part A will consist of 20 MCQ on basic ML concepts and math fundamentals on topics such as probability, statistics and linear algebra. Part B will consist of two Programming questions. The overall test duration will be 75 minutes.

I'll solve each problem step-by-step:

1. Two squares are chosen at random on a chessboard. What is the probability that they have a side in common?

- a) $8/13$
- b) $17/18$
- c) $5/13$
- d) $1/18$

Two squares with a side in common on a chessboard:

There are 64 squares on a chessboard. Let's analyze the favorable cases:

- **Edge squares:** There are 4 squares on each edge (top, bottom, left, right) that can only share a side with a maximum of 3 other squares. $(4 \text{ edges} * 3 \text{ squares/edge}) = 12 \text{ favorable cases.}$

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

- **Inner squares:** There are $4 \times 6 = 24$ squares in the inner grid. Each can share a side with 4 other squares. $(24 \text{ squares} \times 4 \text{ squares/square}) = 96$ favorable cases.

Total favorable cases: $12 \text{ (edge)} + 96 \text{ (inner)} = 108$

The total number of ways to choose two squares is $64C2$ (combinations of 64 elements taken 2 at a time) $= (64 \times 63) / 2 = 2016$

Probability: Favorable cases / Total cases $= 108 / 2016 = 1/18$

Therefore, answer choice (d) 1/18 is correct.

2. The police plans to enforce speed limits during morning rush hour on four different routes into the city. The traps on routes A, B, C, and D are operated 40%, 30%, 20%, and 30% of the time, respectively. Biff always speeds to work, and he has probability 0.2, 0.1, 0.5, and 0.2 of using those routes. What is the probability that he'll get a ticket on any one morning?

- a) 0.27
- b) 0.93
- c) 0.73
- d) 0.07

Biff getting a speeding ticket:

We need to calculate the probability that Biff speeds and encounters a speed trap.

Probability breakdown:

- Biff uses route A: $0.2 \text{ (Biff's probability)} \times 0.4 \text{ (speed trap probability)} = 0.08$
- Biff uses route B: $0.1 \times 0.3 = 0.03$
- Biff uses route C: $0.5 \times 0.2 = 0.1$
- Biff uses route D: $0.2 \times 0.3 = 0.06$

Total probability: $0.08 + 0.03 + 0.1 + 0.06 = 0.27$

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

Therefore, answer choice (a) 0.27 is correct.

The probability that he will receive a speeding ticket by passing through these locations is 0.27.

Question 3: Correlation between X and Y

1. Start with the given information:

$$\text{Var}(X) = 1$$

$$\text{Var}(Y) = 4$$

$$\text{Var}(2X - 3Y) = 34$$

2. Use the formula for the variance of a linear combination of random variables:

$$\text{Var}(aX + bY) = a^2 \cdot \text{Var}(X) + b^2 \cdot \text{Var}(Y) + 2ab \cdot \text{Cov}(X, Y)$$

Plugging in the values:

$$34 = 2^2 \cdot 1 + (-3)^2 \cdot 4 + 2 \cdot 2 \cdot (-3) \cdot \text{Cov}(X, Y)$$

3. Simplify the equation:

$$34 = 4 + 36 - 12 \cdot \text{Cov}(X, Y)$$

4. Solve for $\text{Cov}(X, Y)$:

$$34 - 40 = -12 \cdot \text{Cov}(X, Y)$$

$$-6 = -12 \cdot \text{Cov}(X, Y)$$

$$\text{Cov}(X, Y) = -6 / -12$$

$$\text{Cov}(X, Y) = 1/2$$

5. Use the formula for correlation:

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

$$\rho(X, Y) = \text{Cov}(X, Y) / (\text{sqrt}(\text{Var}(X)) * \text{sqrt}(\text{Var}(Y)))$$

Plugging in the values:

$$\rho(X, Y) = (1/2) / (\text{sqrt}(1) * \text{sqrt}(4))$$

$$\rho(X, Y) = (1/2) / (1 * 2)$$

$$\rho(X, Y) = 1/4$$

So, the correlation between X and Y is 1/4.

4. Probability of picking the fair coin with two heads:

To solve this problem, we can use Bayes' theorem. Let's define the events:

A: Picking the fair coin B: Getting heads twice

We want to find the probability of event A given event B, $P(A|B)$.

According to Bayes' theorem, we have:

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

$P(B|A)$ is the probability of getting heads twice given that we picked the fair coin.

Since the fair coin comes up heads with a probability of 1/2, the probability of getting heads twice is $(1/2)^2 = 1/4$.

$P(A)$ is the probability of picking the fair coin, which is 1/2 since we randomly picked a coin.

$P(B)$ is the probability of getting heads twice. This can be calculated by considering both cases: picking the fair coin and getting heads twice, and picking the biased coin and getting heads twice.

For the biased coin, the probability of getting heads twice is $(3/4)^2 = 9/16$.

So, $P(B) = P(B|A) * P(A) + P(B|\text{not } A) * P(\text{not } A) = (1/4) * (1/2) + (9/16) * (1/2) = 1/8 + 9/32 = 13/32$.

Now, we can substitute these values into Bayes' theorem:

$$P(A|B) = (P(B|A) * P(A)) / P(B) = (1/4) * (1/2) / (13/32) = 1/8 / (13/32) = 1/8 * (32/13) = 4/13.$$

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

Therefore, the probability that you picked the fair coin given that you got heads twice is $\frac{4}{13}$.

So, the correct option is b) $\frac{4}{13}$.

Question 5: PCA

To determine the correct statements about PCA:

- (i) We must standardize the data before applying: This is generally true because PCA is affected by the scale of the data. Standardizing (subtracting the mean and dividing by the standard deviation) ensures that each feature contributes equally to the analysis.
- (ii) We should select the principal components which explain the highest variance: This is true. PCA is designed to find the directions (principal components) that maximize the variance in the data. Therefore, the principal components with the highest variance are the most important.
- (iii) We should select the principal components which explain the lowest variance: This is false. We generally discard components that explain little variance as they do not contribute much to the data's structure.
- (iv) We can use PCA for visualizing the data in lower dimensions: This is true. One of the key applications of PCA is to reduce the dimensionality of the data for visualization, usually in 2D or 3D.

So, the correct answer is:

- a) (i), (ii), and (iv)

Question 6: Number of solutions for the system of equations

The given system of equations is:

$$2x + y - z = 4$$

$$x - 2y + z = -2$$

$$-x + 2y - z = -2$$

.

Step 1: Convert the system of equations to a matrix form

We can represent the system of equations in a matrix form using the coefficient matrix and constant vector. The coefficient matrix A will contain the coefficients of the variables, and the constant vector b will contain the constants on the right side of the equations.

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

For this system, we have:

$$\begin{array}{ccc|ccc} 2 & 1 & -1 & | & x & | & 4 & | \\ 1 & -2 & 1 & | & * & | & y & | & = & | & -2 & | \\ -1 & 2 & -1 & | & & | & z & | & & | & -2 & | \end{array}$$

This can be written mathematically as:

$$A * X = b$$

where:

$$A = \begin{bmatrix} 2 & 1 & -1 \\ 1 & -2 & 1 \\ -1 & 2 & -1 \end{bmatrix} \quad X = [x, y, z] \quad b = [4, -2, -2]$$

Step 2: Analyze the determinant of the coefficient matrix

The determinant of the coefficient matrix ($\det(A)$) plays a crucial role in determining the number of solutions. Here are the possibilities:

- If $\det(A) = 0$: This indicates that the rows of the matrix are linearly dependent. In this case, there are either infinitely many solutions or no solutions, depending on the specific system.
- If $\det(A) \neq 0$: This indicates that the rows of the matrix are linearly independent. In this case, there exists a unique solution for the system of equations.

Step 3: Solve the system using Gaussian elimination (optional)

Gaussian elimination is a technique for solving systems of linear equations. By transforming the augmented matrix (A appended with b) into an upper triangular form, we can back-substitute to solve for the variables.

If $\det(A) = 0$, performing Gaussian elimination can help us determine the following:

- If the system can be reduced to a row of zeros with a non-zero constant on the right side, there are no solutions (inconsistent system).

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

- If the system can be reduced to a row of zeros with a zero constant on the right side, there are infinitely many solutions.

Step 4: Interpret the results based on the determinant

- If $\det(A) = 0$ and Gaussian elimination leads to a row of zeros with a non-zero constant on the right side, there are **a) 0 solutions**. (Inconsistent system)
- If $\det(A) = 0$ and Gaussian elimination leads to a row of zeros with a zero constant on the right side, there are **c) infinitely many solutions**.
- If $\det(A) \neq 0$, there is a unique solution, which is **b) 1**.

Solving the system

Now, let's apply these steps to the given system:

1. We already have the coefficient matrix A and constant vector b .
2. We can calculate the determinant of A using libraries like NumPy in Python or any other mathematical tool. In this case, $\det(A) = 0$.
3. Since $\det(A) = 0$, the system might have either infinitely many solutions or no solutions. Let's try using Gaussian elimination (not shown here, but you can use a calculator or coding tools). If Gaussian elimination results in a row of zeros with a non-zero constant on the right side, there are no solutions. If it results in a row of zeros with a zero constant on the right side, there are infinitely many solutions.

Conclusion

By analyzing the determinant and using Gaussian elimination (if necessary), we can determine that the given system of equations has **a) 0 solutions**. There is an inconsistency in the system that prevents any solution from satisfying all three equations simultaneously.

We have a row of zeros on the left-hand side but a non-zero value on the right-hand side in the third row, indicating an inconsistency.

So, the system of equations has no solution:

- a) 0

Question 7: Output of a 3-input neuron

The given information includes:

- Weights: 1, 4, 3
- Inputs: 4, 8, 5
- Transfer function is linear with a constant of proportionality of 3

First, calculate the weighted sum of the inputs:

$$\text{Weighted sum} = (1 \times 4) + (4 \times 8) + (3 \times 5)$$

$$\text{Weighted sum} = 4 + 32 + 15$$

$$\text{Weighted sum} = 51$$

Since the transfer function is linear with a proportionality constant of 3, the output will be:

$$\text{Output} = 3 \times \text{Weighted sum}$$

$$\text{Output} = 3 \times 51$$

$$\text{Output} = 153$$

So, the correct answer is:

b) 153

Question 8: Impact on p-value in Medical Screening

In medical screening, to avoid false negatives, we want to ensure that we detect a disease if it is present. This means we want to increase the sensitivity of the test. Sensitivity is the true positive rate, and to increase it, we might accept a higher false positive rate. The p-value is the threshold at which we reject the null hypothesis (no disease).

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

- Choosing a higher p-value (a): This means we are more willing to reject the null hypothesis, which increases the chance of detecting the disease (true positive), but also increases the chance of false positives.
- Choosing a lower p-value (b): This would make the test more stringent, decreasing false positives but potentially increasing false negatives, which is not what we want.
- Choosing the same p-value (c): This would not address the concern of minimizing false negatives.
- False negatives do not depend on p-value (d): False negatives are indeed related to the choice of p-value as it affects the sensitivity and specificity of the test.

Therefore, to minimize false negatives, we would choose a higher p-value:

- a) we would choose a higher p-value

Question 9: Upper Bound of the Likelihood in Logistic Classification

In logistic classification, the likelihood function measures how well the model's predicted probabilities match the actual outcomes. The likelihood for a correctly predicted outcome is given by the predicted probability p , and for logistic regression, this probability ranges from 0 to 1.

- **a) 1, Yes:** The value 1 represents a perfect prediction where the probability of the predicted class is 1. This value is attainable in theory for the likelihood of a single observation.
- **b) e, No:** The number e (approximately 2.718) is related to the natural logarithm base but not directly relevant to the likelihood's range in logistic regression.
- **c) 1, No:** While the likelihood of a single observation can be 1, the combined likelihood of all observations being 1 simultaneously is not practically attainable in most real-world scenarios due to model and data limitations.
- **d) 0, Yes:** A likelihood of 0 means a complete mismatch between predictions and actual outcomes, which is not desirable.

The upper bound of the likelihood for a single observation is 1, but for the combined likelihood of all observations, achieving 1 is generally not feasible.

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

- c) 1, No

Question 10: Mean and Variance Estimators

Step-by-Step Analysis of Mean and Variance Estimators:

We'll analyze each statement and identify the correct answer based on the properties of unbiased estimators.

Statement 1:

- **S^2 is an unbiased estimator of σ^2 :** This statement is **true**. Here's why:

The sample variance (S^2) is calculated as the sum of squared deviations from the mean divided by $n-1$ (n is the sample size). This formula effectively estimates the average squared deviations from the population mean in a finite sample. It can be proven mathematically that $E(S^2) = \sigma^2$, where E denotes expected value. This implies S^2 is an unbiased estimator of σ^2 .

- **S is an unbiased estimator of σ :** This statement is **false**. Here's why:

The sample mean (S) is an unbiased estimator of the population mean (μ). However, taking the square root of an unbiased estimator does not necessarily result in another unbiased estimator. In this case, $E(\sqrt{S}) \neq \sqrt{\mu}$ due to a phenomenon called bias amplification.

Statement 2:

- **$(n-1/n) * M$ is an unbiased estimator of μ :** This statement is **false**. Here's why:

The sample mean (M) itself is already an unbiased estimator of the population mean (μ). Scaling it by $(n-1/n)$ does not change its unbiased

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

property. However, it's unnecessary and usually omitted.

- **$(n-1/n) * S^2$ is an unbiased estimator of σ^2 :** This statement is **true**.

Here's why:

While S^2 is an unbiased estimator of σ^2 , it suffers from a slight downward bias for small sample sizes. Dividing S^2 by $(n-1/n)$ corrects for this bias, making $(n-1/n) * S^2$ a more accurate estimator of σ^2 , especially for smaller samples.

Conclusion:

Based on the analysis, only Statement 1 regarding S^2 being an unbiased estimator of σ^2 is true. Therefore, the correct answer is **a) 1 only**.

Question 11: Rank of Matrices

Analysis:

1. We know the property: $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$
2. We are given: $\text{rank}(A) = 2$ and $\text{rank}(AB) = 3$

From property 1, $\text{rank}(AB)$ can't be greater than the minimum rank of A and B. Since $\text{rank}(AB) = 3$ and $\text{rank}(A) = 2$, $\text{rank}(B)$ must be greater than or equal to 3. This eliminates options (a) and (b).

Therefore, the correct answer is: **c) $\text{rank}(B) \geq 3$**

Question 12: Infinite Solutions for System of Equations

Solution Steps:

1. **Write the augmented matrix:**

$$\begin{bmatrix} 1 & 1 & 1 & | & 1 \end{bmatrix}$$

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

$$[a \ -a \ 3 \ | \ 5 - a]$$

$$[5 \ -3 \ a \ | \ 6]$$

2. Perform row operations for echelon form:

- $R2 = R2 - a \cdot R1$:

$$[1 \ 1 \ 1 \ | \ 1]$$

$$[0 \ -2a \ 3-a \ | \ 5 - a]$$

$$[5 \ -3 \ a \ | \ 6]$$

- $R3 = R3 - 5 \cdot R1$:

$$[1 \ 1 \ 1 \ | \ 1]$$

$$[0 \ -2a \ 3-a \ | \ 5 - a]$$

$$[0 \ -8 \ a-5 \ | \ 1]$$

3. Analyze for infinite solutions:

For infinite solutions, the second and third rows must be proportional. Set up the proportion for corresponding elements:

$$-2a / -8 = (3-a) / (a-5) = (5-a) / 1$$

Solve this system of equations to find a :

- Simplify: $a / 4 = (3-a) / (a-5)$
- Cross-multiply and solve the quadratic:

$$a(a-5) = 4(3-a)$$

$$a^2 - 5a = 12 - 4a$$

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

$$a^2 - a - 12 = 0$$

$$(a-4)(a+3) = 0$$

$$a = 4 \text{ or } a = -3$$

Answer:

Therefore, the system has infinite solutions when a is either **a) -3** or **d) 4**.

Question 13: Gradient Descent vs. Normal Equation

Analysis:

Given:

- $m=50$ examples
- $n=200000$ features
- The normal equation involves calculating the inverse of a very large matrix ($n \times n$) when there are many features ($n = 200000$). This is computationally expensive and impractical.
- Gradient descent is an iterative approach that doesn't require matrix inversion, making it more suitable for large datasets with many features.

Therefore, the correct answer is: **a) Gradient descent, since inverse(XTX) will be very slow to compute in the normal equation.**

Question 14: Eigenvalues of a Matrix

Given:

- A 4x4 square matrix with 0's as diagonal elements and 1's as off-diagonal elements.

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

To find the eigenvalues of this matrix, we can use the fact that the sum of the elements in each row (or column) of the matrix gives the eigenvalue.

The given matrix can be represented as:

[0 1 1 1]

[1 0 1 1]

[1 1 0 1]

[1 1 1 0]

Each row (or column) has the sum 2, which means the eigenvalues are 2 and 0. Since it's a 4x4 matrix, each eigenvalue has a multiplicity of 2.

Thus, the correct answer is:

- d) 4, 0, 0, 0

You're right, relying on external libraries like NumPy might not be ideal during an exam. Here's how you can solve these problems by hand:

15. Eigenvalues of A^{19} (without libraries):

1. **Characteristic Equation:** Find the characteristic equation of the matrix A.

The characteristic equation is $\det(A - \lambda I) = 0$, where A is the matrix, λ is the eigenvalue, and I is the identity matrix.

For $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, the characteristic equation becomes:

$$(1 - \lambda)(-1 - \lambda) - 1 * 1 = \lambda^2 + \lambda - 2 = 0$$

2. **Solve for Eigenvalues:** Solve the characteristic equation for λ . In this case, $\lambda = 1$ or $\lambda = -2$.

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

3. **Raise to Power:** Since we need the eigenvalues of A^{19} , simply raise the obtained eigenvalues (1 and -2) to the power of 19.

Eigenvalues of $A^{19} \approx 1^{19}$ (which is 1) and $(-2)^{19}$ (which is a very large negative number, approximately -512,000).

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$A - \lambda I \Rightarrow \begin{vmatrix} 1-\lambda & 1 \\ 1 & -1-\lambda \end{vmatrix}$$

$$\Rightarrow (1-\lambda)(-1-\lambda) - 1$$

$$\Rightarrow -1 - \cancel{\lambda} + \cancel{\lambda} + \lambda^2 - 1$$

$$\Rightarrow \lambda^2 = 2$$

$$\Rightarrow \lambda = \pm\sqrt{2}.$$

$$A^{19} = \cdot \lambda^{19} \Rightarrow (-\sqrt{2})^{19} = -512\sqrt{2}$$

$$\cdot (+\sqrt{2})^{19} = 512\sqrt{2}.$$

Note: Due to the nature of exponentiation, raising a negative number to an odd power results in a negative answer. However, on exams, focusing on the magnitude (absolute value) might be sufficient. So, the answer choices with the closest magnitudes are most likely the correct ones.

16. Derivative of $y = x^x$: This problem requires logarithmic differentiation. Here's the manual approach:

1. **Logarithm both sides:** Take the equation $y = x^x$ and put both sides of the equation within a $\ln(x)$ function:

$$\ln(y) = \ln(x^x)$$

2. **Differentiate both sides:** Differentiate both sides of the equation using the chain rule. Remember, the derivative of $\ln(x)$ is $1/x$.

$$d/dx(\ln(y)) = d/dx(\ln(x^x)) \quad 1/y * dy/dx = x^{(x-1)} * (1 + \ln(x))$$

3. **Isolate dy/dx :** Rearrange the equation to isolate dy/dx :

$$dy/dx = y * x^{(x-1)} * (1 + \ln(x))$$

4. **Evaluate at $x = 2$:** Substitute $x = 2$ into the equation. However, solving for y directly might be difficult. Notice that $y = 2^2$ (based on the original equation). You can use this to simplify the expression.

Alternative approach for evaluation:

Instead of solving for y , substitute $y = x^x$ into the equation:

$$(x^x) * x^{(x-1)} * (1 + \ln(x)) = dy/dx$$

This simplifies the expression as $x^2 x^{(x-1)}$ becomes $x^{(2x-1)}$. However, evaluating this at $x = 2$ might still be cumbersome.

The answer to question 16 is **b) $4(1+\log 2)$** .

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

We arrived at this answer by applying logarithmic differentiation to find the expression for dy/dx and then evaluating it at $x = 2$. However, directly solving for y at $x = 2$ might be difficult. Therefore, we presented the simplified form $dy/dx = y * (\ln(x) + 1)$ and then noted that for $x = 2$ (where $y = 4$), the expression becomes $4(1+\log 2)$.

Question 17: Characteristic Equation and Matrix Inverse

Given: Characteristic equation of matrix A is $t^2 - t - 1 = 0$

1. Characteristic Equation and Eigenvalues:

The characteristic equation of a matrix A is a polynomial equation in terms of a variable λ , where λ represents the eigenvalues of A . In this case, the characteristic equation is:

$$t^2 - t - 1 = 0$$

We need to solve this equation to find the eigenvalues (λ).

2. Solve for Eigenvalues:

Using the quadratic formula, we can solve the characteristic equation:

$$\lambda = (1 \pm \sqrt{1 + 4 * 1}) / 2$$

$$\lambda = (1 \pm \sqrt{5}) / 2$$

This equation tells us that A has two eigenvalues: $\lambda_1 = (1 + \sqrt{5}) / 2$ and $\lambda_2 = (1 - \sqrt{5}) / 2$.

3. Inverse of a Matrix and Eigenvalues:

For a matrix A to be invertible (have an inverse), its eigenvalues cannot be zero. If an eigenvalue is zero, the determinant of the matrix (which can be obtained from the characteristic equation) will also be zero. An inverse cannot exist for a matrix with a determinant of zero.

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

In our case, both eigenvalues (λ_1 and λ_2) are non-zero. This is a good sign!

4. Conclusion based on Eigenvalues:

Since the characteristic equation of A has no zero eigenvalues, the matrix A likely has an inverse (A^{-1}). However, the characteristic equation alone doesn't provide enough information to determine the exact form of the inverse (like $A - I$ or $A + I$).

Therefore, the answer is b) A^{-1} exists but cannot be determined from the data.

Question 18: Bias and Variance of Different Models

Given:

- Fitting data from a cubic function corrupted by standard Gaussian noise using linear and 5th-degree polynomial models.

Options:

- a) $\text{Bias}(M_1) \leq \text{Bias}(M_5)$, $\text{Variance}(M_1) \leq \text{Variance}(M_5)$
- b) $\text{Bias}(M_1) \geq \text{Bias}(M_5)$, $\text{Variance}(M_1) \leq \text{Variance}(M_5)$
- c) $\text{Bias}(M_1) \leq \text{Bias}(M_5)$, $\text{Variance}(M_1) \geq \text{Variance}(M_5)$
- d) $\text{Bias}(M_1) \geq \text{Bias}(M_5)$, $\text{Variance}(M_1) \geq \text{Variance}(M_5)$

Explanation:

- A 5th-degree polynomial model (M_5) is more complex than a linear model (M_1), so it has the potential to capture more intricate patterns in the data, leading to lower bias. Therefore, $\text{Bias}(M_1) \geq \text{Bias}(M_5)$
- However, a more complex model tends to have higher variance because it may fit noise in the data. Therefore, $\text{Variance}(M_1) \leq \text{Variance}(M_5)$

Follow Vikram Gaur : <https://www.linkedin.com/in/vikram-gaur-0252aa185>

Thus, the correct answer is:

- b) $\text{Bias}(M_1) \geq \text{Bias}(M_5)$, $\text{Variance}(M_1) \leq \text{Variance}(M_5)$