

Editorial: Mock: End Module B Quiz → June 21

Easy Level Questions (15)

1. Regression

Type: MCQ

Story: Sarah operates "Sunny Lemonade," a popular stand in Mumbai. She's been tracking daily sales against temperature for 3 months. After analyzing the data, her data scientist friend Dev creates this predictive model:

$$\text{Sales (₹)} = 10 \times \text{Temperature (°C)} + 50$$

One scorching afternoon, the weather app shows 25°C - a typical May day. Sarah wants to prepare the right amount of lemonade. What sales should she expect according to Dev's model?

- A) ₹250 (underprepares by 50%)
- B) ₹300 (exact prediction)
- C) ₹350 (overprepares by 16.6%)
- D) ₹400 (risks wastage)

Answer: B) ₹300

Explanation:

Breaking down the calculation:

1. Temperature effect: $10 \times 25^\circ\text{C} = ₹250$
2. Base sales: + ₹50 (from factors like location popularity)
3. Total expected: $₹250 + ₹50 = ₹300$

Business Impact:

- Choosing ₹300 helps Sarah:
 - ✓ Optimize lemonade batch size
 - ✓ Minimize ingredient waste
 - ✓ Meet customer demand precisely

Why other options fail:

- A) ₹250 → Runs out by afternoon
- C/D) Overproduction → Spoiled inventory

2. Polynomial Regression

Type: MCQ

Story: During a physics experiment, Maya launches a ball into the air. Her sensor records the

height (in meters) over time (seconds) following the equation:

$$h = -2t^2 + 10t + 5$$

where:

- $-2t^2$ represents gravity's effect
- $10t$ is the initial velocity
- 5 is the launch height

At the 2-second mark, Maya's display malfunctions. What height should the sensor have shown?

- A) 15m
- B) 17m
- C) 19m
- D) 21m

Answer: B) 17m

Explanation:

Substituting $t = 2$:

1. Gravity effect: $-2 \cdot (2)^2 = -8\text{m}$
2. Velocity contribution: $10 \cdot 2 = +20\text{m}$
3. Launch height: $+5\text{m}$
Total height = $-8 + 20 + 5 = 17\text{m}$

Why not others?

- A) 15m ignores the launch height
 - C) 19m miscalculates gravity's effect
 - D) 21m doubles the velocity term
-

3. Dimensionality Reduction

Type: MCQ

Story: Dr. Lee at General Hospital is analyzing patient health records with 100+ measurements (blood pressure, cholesterol, etc.). His team struggles to spot patterns in such high-dimensional data. He wants to visualize the data in 2D while keeping the most important information. Which solution should he choose?

- A) Linear Regression (predicts outcomes but doesn't reduce dimensions)
- B) PCA (can project data to lower dimensions)
- C) K-Means (groups data but doesn't visualize it)

- D) SVM (classifies data but maintains original dimensions)

Answer: B) PCA

Explanation: PCA identifies key patterns by transforming data into principal components, allowing 2D visualization while preserving 85-90% of variance in typical medical datasets.

Why not others?

- A) Predicts but doesn't reduce
 - C) Clusters but keeps original features
 - D) Creates boundaries but doesn't compress data
-

4. Page Rank

Type: MCQ

Story: In a small academic wiki, Professor X writes a page about "Neural Networks" that only links to Professor Y's "Deep Learning" page. Professor Y's page only links back to Professor X's page. The university web crawler is calculating initial importance scores before considering links. What starting PageRank does Professor X's page receive?

- A) 0.25 (like 4 pages exist)
- B) 0.50 (correct for 2-page system)
- C) 0.75 (would imply bias)
- D) 1.00 (all importance to one page)

Answer: B) 0.50

Explanation: PageRank initializes with uniform distribution: $1/N$ where $N=2$ pages (X and Y). This ensures fairness before link analysis begins.

Real-world implication:

This models how Google treated all new webpages equally before analyzing their connections.

5. Foundations of the Perceptron

Type: MCQ

Story: At FreshFruit Sorting Co., engineer Priya is building an automated classifier to separate apples and oranges on a conveyor belt. The system uses:

- **Weight** (in grams)
- **Color** (1 = red, 0 = orange)

Her perceptron model has:

- Weights: [0.02 (for weight), -1 (for color)]
- Bias: 0.5

One fruit arrives with:

- Weight = 150g
- Color sensor reads 1 (red)

The control system needs to decide: should this fruit be routed to the apples bin (1) or oranges bin (0)?

What's the correct classification?

- A) Apple
- B) Orange

Answer: A) Apple

Explanation:

1. **Weight contribution:** $0.02 \times 150 = 3.0$
(Heavier fruits lean toward apple classification)
2. **Color penalty:** $-1 \times 1 = -1.0$
(Red color strongly suggests apple, but the negative weight inverts this)
3. **Bias adjustment:** $+0.5$
(System's inherent preference for apples)
4. **Decision:** $3.0 - 1.0 + 0.5 = 2.5 > 0 \rightarrow \text{Apple}$

Real-World Implications:

- The negative color weight seems counterintuitive but could account for:
 - Some apples being green (color=0)
 - Rare red oranges (false positives)
- The heavy weight coefficient shows mass is the dominant factor

Why not Orange?

The combined evidence (heavy weight + color) outweighs the system's initial bias, despite the color weight paradox.

6. Enhancing the Perceptron with Learning Algorithms

Type: MCQ

Story: A perceptron misclassifies a point. What happens next?

- A) Weights adjust
- B) Bias becomes zero

- C) Learning rate increases
- D) Nothing

Answer: A) Weights adjust

Explanation: Learning rule updates weights after misclassification.

7. Neural Networks - Forward Propagation

Type: MCQ

Story: MediScan AI is developing a neural network to analyze X-ray images. In one test case, their network's first neuron receives an input value of -3 (representing an unusual bone density reading). The neuron uses ReLU activation. What value does it pass to the next layer?

- A) -3 (passes negative value)
- B) 0 (ReLU's output)
- C) 3 (absolute value)
- D) 1 (default value)

Answer: B) 0

Explanation:

The ReLU (Rectified Linear Unit) activation works like a medical filter:

- Input: -3 (concerning but possibly erroneous reading)
- Operation: $\max(0, -3) = 0$
- Output: 0 (flags as "normal" for further analysis)

Clinical Implications:

- Prevents negative values from distorting diagnosis
- Allows the network to focus on medically significant positive findings
- Similar to how radiologists ignore certain artifacts

Why other options could be dangerous:

- A) Could amplify imaging errors
 - C) Might create false positives
 - D) Would lose all diagnostic information
-

8. Regularization

Amit is working on building a regression model for predicting house prices. He has a large number of features (variables) in his dataset. However, Amit does not want automatic

feature selection to occur in his model because he believes that all the features provide valuable information. Given Amit's preference to avoid automatic feature selection, which regularization method should he choose?

- A. L1 Regularization (Lasso)
- B. L2 Regularization (Ridge)

Answer: B. L2 Regularization (Ridge)

Explanation:

In general, when a weight W_i has already been small in magnitude, L2 does not care to reduce it to zero, L2 would rather reduce big weights than eliminate small weights to 0. The result is that the weights are reduced, but almost never reduced to 0, i.e. almost never be completely eliminated, meaning no feature selection. On the other hand, L1 cares about reducing big weights and small weights equally. For L1, the less informative features get reduced. Some features may get completely eliminated by L1, thus we have feature selection.

9. SVM

Type: MCQ

Story: An SVM classifies cats vs. dogs. What does a high C value imply?

- A) More margin violations allowed
- B) Strict margin enforcement
- C) Ignores outliers
- D) Reduces dimensionality

Answer: B) Strict margin enforcement

Explanation: High C penalizes misclassifications heavily.

10. Logistic Regression

Which of the following statements about Logistic Regression is TRUE?

- A. Logistic Regression can be used for regression problems.
- B. The odds in Logistic Regression are defined as $p / (1 - p)$, where p is the probability of the positive class.
- C. The log-odds (logit) in Logistic Regression are given by $\ln(p / (1 - p))$.
- D. Logistic Regression outputs probabilities that can be converted into class labels.
- E. The probability in Logistic Regression can be expressed as $P = 1 / (1 + e^{-z})$, where z is the linear combination of the input features.
- F. Logistic Regression is a unsupervised ML method.

Answer: The correct answers are B, C, D, and E. **Explanation:**

- Odds represent the ratio of the probability of the positive class to the probability of the negative class.
- The logit function is the natural logarithm of the odds.
- Logistic Regression models the probability of the positive class and these probabilities can be converted into class labels using a threshold.
- $P = 1 / (1 + e^{-z})$, where z is the linear combination of the input features. This is the sigmoid function that maps any real-valued number into the $[0, 1]$ interval.
- Logistic Regression is used for classification problems, not regression.
- Logistic Regression is a Supervised ML method as it needs Y labels.

11. Unsupervised Learning - K Means Clustering

Type: MCQ

Story: MegaMall is analyzing customer spending habits using transaction data (purchase frequency, average spend). Their data scientist chooses K-means to group customers into 3 segments. What core calculation drives the clustering algorithm?

- A) Euclidean distance (measures straight-line distance between customers and cluster centers)
- B) R-squared (evaluates linear regression fit)
- C) P-value (tests statistical significance)
- D) Silhouette score (validates cluster quality)

Answer: A) Euclidean distance

Explanation:

- **How it works:**
 - Calculates distance from each customer to all 3 cluster centroids
 - Assigns customer to nearest cluster
 - Repeats until clusters stabilize
- **Business Impact:**
 - Groups similar spenders together
 - Enables targeted marketing campaigns
 - Identifies high-value customer segments

Real-world Example:

Customers clustered by distance might reveal:

- Cluster 1: Occasional big spenders (far from others)

- Cluster 2: Frequent small spenders
 - Cluster 3: Average weekly shoppers
-

12. Unsupervised Learning - Evaluation Metrics

Type: MCQ

Story: After running customer clustering, MegaMall's team gets a Silhouette score of 0.2. The marketing manager asks what this means for their "Targeted Discounts" campaign.

- A) "Strong clustering" (clear segments for personalized offers)
- B) "Weak clustering" (customer groups overlap significantly)
- C) "Perfect separation" (ideal for 1:1 marketing)
- D) "Random guessing" (no patterns detected)

Answer: B) Weak clustering

Explanation:

- **Score Interpretation:**
 - +1 = Excellent separation
 - 0 = Overlapping clusters
 - -1 = Incorrect grouping
 - **0.2 Implication:**
 - Some cluster structure exists
 - But 30-40% of customers could fit multiple groups
 - **Campaign Adjustment Needed:**
 - ✓ Broaden discount categories
 - ✓ Use additional demographic data
 - ✓ Consider fewer clusters
-

13. Regression

Type: NAT

Story: CycleCity's rental prediction model for their smart bikes is:

$$\text{Rentals} = 5 \times \text{Temperature} + 20$$

Where:

- Rentals = Hourly bike rentals
- Temperature = °C at the station

On a cool morning (10°C), how many rentals should the system prepare for? (Integer answer)

Answer: 70

Calculation:

1. Temperature effect: $5 \times 10^\circ\text{C} = 50$ rentals
2. Base demand: + 20 rentals
3. Total: $50 + 20 = 70$ rentals

Operational Impact:

- Ensures 70 bikes are available/charged
- Staffs maintenance crews appropriately
- Balances redistribution trucks

Model Limitations:

- Doesn't account for rain/holidays
 - Base 20 rentals may vary by neighborhood
-

14. Polynomial Regression

Type: NAT

Story: SpaceX's test flight team is modeling their experimental rocket's altitude (in km) using:

$$h(t) = t^3 - 2t + 10$$

where:

- t = flight time (minutes)
- t^3 accounts for increasing thrust
- $-2t$ represents fuel burn effects

At $t=3$ minutes during re-entry, the telemetry display fails. What altitude should engineers expect? (Integer answer)

Answer: 31

Calculation:

1. Thrust effect: $3^3 = 27\text{km}$
2. Fuel penalty: $-2 \times 3 = -6\text{km}$
3. Launch altitude: $+10\text{km}$
Total: $27 - 6 + 10 = 31\text{km}$

Emergency Protocol:

- Triggers parachute deployment if $h < 30\text{km}$

- Activates backup comms if $h > 35\text{km}$
 - 31km means:
 - ✓ Normal descent trajectory
 - ✓ Continue primary recovery plan
-

15. Page Rank

Type: NAT

Story: Wikipedia's new "Space Exploration" category has 3 circularly linked pages:

1. Moon Landings → Mars Missions
2. Mars Missions → Space Stations
3. Space Stations → Moon Landings

The algorithm initializes with equal ranks. After one update (damping=0.85), what's the Moon Landings page's rank? (Answer $\times 100$). Round to nearest integer.

Answer: 33

Step-by-Step:

1. Initial PR = $1/3 \approx 0.33$ for all
 2. $PR(\text{Moon}) = (1-0.85)/3 + 0.85 \times (PR(\text{Space Stations})/1)$
 3. $= 0.05 + 0.85 \times 0.33 \approx 0.3305 \rightarrow 33.05$
-

16. Dimensionality Reduction

Type: MSQ

Story: NASA's Mars rover has 1000+ sensors generating terrabyte-scale data. Dr. Chen must simplify this for Earth transmission while preserving key geological patterns. Which PCA steps are essential? (Select all)

- A) Center the data (shifts features to zero-mean)
- B) Compute covariance matrix (finds feature relationships)
- C) Use labels to guide projection (PCA is unsupervised)
- D) Select top eigenvectors (keeps directions of max variance)

Answers: A, B, D

Why These Matter:

1. **Centering:** Enables comparing sensors on same scale
2. **Covariance:** Reveals which sensors vary together

3. **Eigenvectors:** Identifies "main directions" of Mars terrain

Transmission Savings:

- Reduces 1000 sensors → 20 principal components
- Maintains 95% variance with 98% less data

Critical Note:

PCA works without labels (C is wrong) - it finds natural patterns, not human categories

17. Neural Networks - Backward Propagation

Type: NAT

Story: DeepMind Robotics is training a neural arm to grasp objects. During backpropagation, the system calculates a gradient of -0.1 for a critical weight controlling grip pressure. With a conservative learning rate of 0.01, what absolute adjustment should be made to this weight? (Multiply answer by 1000)

Answer: 1

Technical Breakdown:

1. **Gradient Interpretation:**

- Negative sign indicates reducing the weight improves performance
- Magnitude 0.1 suggests moderate adjustment needed

2. **Update Calculation:**

$$\Delta \text{weight} = - (\text{learning rate} \times \text{gradient}) = - (0.01 \times -0.1) = +0.001$$

$$\text{Absolute value} \times 100 = 0.001 \times 1000 = 1$$

Safety Implications:

- This tiny adjustment (0.001):
 - ✓ Prevents overcompensation in delicate grip control
 - ✓ Allows 100+ iterations for gradual optimization
 - ✓ Maintains stable operation during continuous learning
-

18. CNN

Type: MCQ

Story: Tesla's vision team is processing 7x7 pixel road sign patches through a convolutional layer with:

- 3x3 edge-detection kernel
- Stride=2 (for faster processing)

- Padding=1 (to preserve border features)

The hardware accelerator needs to pre-allocate memory for the output. What dimensions should engineers specify?

- A) 3x3 (insufficient for padded image)
- B) 4x4 (correct output)
- C) 5x5 (would require different stride)
- D) 6x6 (only possible with zero padding)

Answer: B) 4x4

Architecture Insight:

1. **Padded Size:** $7 + 2 \times 1 = 9$ (adds 1-pixel border)
2. **Convolution Math:**
Output = $\lfloor (9 - 3) / 2 \rfloor + 1 = \lfloor 3 \rfloor + 1 = 4$

Real-Time Impact:

- 4x4 output means:
 - ✓ 67% reduction from original size
 - ✓ Fits perfectly in their FPGA memory blocks
 - ✓ Maintains critical edge information for sign recognition

Why Other Sizes Fail:

- A) Loses 25% of sign features
- C/D) Would violate Tesla's hardware constraints

19. SVM

Type: MSQ

Story: Boston Dynamics' new robot uses SVM to classify terrain (grass/concrete/gravel). The RBF kernel model struggles with mixed surfaces. Which tuning approaches would help? (Select all)

- A) Increase C from 1 to 10 (tolerate fewer misclassifications)
- B) Adjust gamma from 0.1 to 0.5 (sharper decision boundaries)
- C) Switch to linear kernel (would fail on non-linear terrain)
- D) Use less training data (worsens performance)

Answers: A, B

Engineering Rationale:

1. **Higher C (10):**

- Prioritizes perfect classification on training data
- Essential for safety-critical terrain detection

2. Tuned gamma (0.5):

- Makes the model more sensitive to local texture changes
- Better captures gravel/grass transitions

Field Test Results:

After tuning:

- 92% accuracy vs. original 85%
- False positives reduced from 8% to 3%
- No more "mixed surface hesitation" incidents

20. Supervised Learning - Evaluation Metrics

Type: NAT

Story: At Mayo Clinic's AI Lab, researchers are testing a new deep learning system for detecting early-stage pancreatic cancer. In clinical trials with 100 high-risk patients (50 confirmed cases via biopsy):

- **Overall Accuracy:** 80%
- **Recall (Sensitivity):** 90%

The hospital board demands to know: *How many true positive cases did the system correctly identify before approving its deployment?*

Answer: 45

Diagnostic Breakdown:

1. Recall Formula:

$$\text{True Positives} / (\text{True Positives} + \text{False Negatives}) = 90\%$$

2. Given:

Total diseased patients = 50

3. Calculation:

$$\text{TP} / 50 = 0.9 \rightarrow \text{TP} = 45$$

Clinical Implications:

- **45 patients** correctly diagnosed early:
 - ✓ Eligible for life-saving interventions
 - ✓ 90% detection rate exceeds human radiologists (avg. 82%)

- **5 false negatives:**
X System missed 5 cases (urgently needs improvement)
- **Accuracy Context:**
80% seems good, but misleading - high recall is critical for fatal diseases

Why This Matters:

- Each 1% recall improvement → Saves ~1,500 lives annually (US estimates)
- FDA requires minimum 85% recall for cancer detection tools

21. Foundations of the Perceptron

Type: NAT

Story: RoboGrocer's smart checkout uses a perceptron to classify fruits. For mangoes:

- **Features:** [Size=1 (large), Blemishes=2 (minor)]
- **Weights:** [2 (size importance), -1 (blemish penalty)]
- **Bias:** 1 (base premium probability)

What weighted sum determines if this mango gets premium pricing? (Integer)

Answer: 1

Calculation:

$$(2 \times 1) + (-1 \times 2) + 1 = 1$$

Business Impact:

- Sum > 0 → ₹200 premium price
- Sum ≤ 0 → ₹150 standard price
- This mango barely qualifies as premium

22. Enhancing the Perceptron with Learning Algorithms

Type: MCQ

Story: Bengaluru's e-scooter startup trains a perceptron to detect pedestrians. At night, it fails 30% of the time. The perceptron convergence theorem suggests:

- A) **Linearly separable** night/day data guarantees improvement
- B) Zero learning rate stops all updates
- C) Higher resolution cameras add complexity
- D) Shuffled data breaks time patterns

Answer: A) Linearly separable night/day data guarantees improvement

Solution: Engineers add infrared features to make night/day patterns separable.

23. Neural Networks - Backward Propagation

Type: NAT

Story: A neural network diagnosing lung cancer calculates:

- Gradient: -0.2 (suggests reducing weight)
- Learning rate: 0.1 (conservative for medical use)

What's the absolute weight update? (×100)

Answer: 2

Math: $0.1 \times 0.2 = 0.02 \rightarrow 2$

Safety Protocol:

- Updates >5 require doctor approval
 - 2 is within safe auto-update range
-

24. CNN

Type: MCQ

Story: AIIMS processes 5×5 MRI slices through:

- 2×2 tumor-detection kernel
- Stride=1, no padding

What output size allocates GPU memory correctly?

- A) 4×4 (correct)
- B) 5×5 (input size)
- C) 6×6 (requires padding)
- D) 3×3 (too aggressive)

Answer: A) 4×4

Formula: $(5 - 2 + 1) = 4$

25. SVM - Optimizing Fraud Detection (MSQ)

Scenario:

ICICI Bank's AI team is optimizing their **SVM-based fraud detection system**. They currently use three different **kernels** to classify transactions:

- **A) RBF Kernel** → Captures **complex non-linear transaction patterns**.
- **B) Polynomial Kernel** → Helps model **amount-frequency relationships**.

- **C) Linear Kernel** → Used for **rapid processing** of clearly separable cases.

Objective:

The goal is to **reduce false positives** (legitimate transactions wrongly flagged as fraud).

Question:

Which of the following **kernel combinations** effectively reduce false positives? (*Select all that apply*)

Options:

- **A) RBF only**
- **B) Polynomial only**
- **C) Linear only**
- **D) RBF + Polynomial**
- **E) RBF + Linear**
- **F) Polynomial + Linear**
- **G) All three (RBF + Polynomial + Linear)**

Solution Explanation:

- **RBF Kernel:** Excellent for **complex, non-linear fraud patterns**, but may lead to **overfitting** if used alone.
- **Polynomial Kernel:** Captures **frequency-based fraud patterns**, but **higher degrees** may cause **excessive false positives**.
- **Linear Kernel:** Fast and effective for **obvious fraud cases**, but **struggles** with **intricate fraud behaviors**.

Best Approach to Reduce False Positives:

✓ **D) RBF + Polynomial** → Balances **pattern recognition** and **structured fraud behavior analysis**.

✓ **F) Polynomial + Linear** → Helps classify structured fraud cases with better precision.

✓ **G) All three kernels** → Provides the most **comprehensive fraud detection model** with a balance of **speed, accuracy, and adaptability**.

Thus, the **best selections** to reduce false positives are:

✓ **D, F, and G**

26. Supervised Learning - Evaluation Metrics

Type: NAT

Story: Gmail's spam filter shows:

- Precision: 0.6 (60% flagged are spam)
- Recall: 0.4 (catches 40% of spam)

Calculate F1-score ($\times 100$):

Answer: 48

Formula: $2 \times (0.6 \times 0.4) / (0.6 + 0.4) = 0.48$

27. Unsupervised Learning - Evaluation Metrics

Type: MCQ

Story:

Flipkart clusters **10,000 customers** and evaluates the clustering performance using:

- **Davies-Bouldin Index:** 0.3 (lower is better, indicating well-separated clusters)
- **Silhouette Score:** 0.15 (closer to 1 is better, indicating poor cluster cohesion)

Given these metrics, what should the **CMO** prioritize?

- A)** Trust the good Davies-Bouldin index
 - B)** Focus on the low Silhouette score
 - C)** Increase the number of clusters
 - D)** Restart the process with PCA
-

Solution Explanation:

- **Davies-Bouldin Index (DBI) = 0.3**
 - A **lower** DBI (close to **0**) is **good**, suggesting that clusters are **well-separated**.
- **Silhouette Score = 0.15**
 - A **low** score (close to **0**) indicates **poor cohesion**, meaning customers **may not belong clearly to a single cluster**.
- **Key Decision:**
 - The **Silhouette Score** suggests weak clustering, meaning **the CMO should focus on improving cluster quality**.

Correct Answer:

✓ **B) Focus on the low Silhouette score**

28. Neural Networks - Backward Propagation

Type: NAT

Story: NeuroTech is training a robotic arm to perform microsurgery. The neural controller has:

- Current weights: [0.5 (for force), -0.5 (for precision)]
- Input from sensors: [2 (tissue resistance), 3 (tremor level)]
- Gradient: -0.2 (indicating excessive force)

With a conservative learning rate (0.1) to ensure patient safety, what should be the updated force control weight? (Round to 1 decimal)

- A) 0.4
- B) 0.5
- C) 0.54
- D) 0.6

Solution

Weight Update Formula:

The weight update in gradient descent follows:

$$[w_{\text{new}} = w_{\text{old}} - \eta \times \text{gradient} \times \text{input}]$$

Substituting values:

$$[w_{\text{new}} = 0.5 - (0.1 \times -0.2 \times 2)]$$

$$[w_{\text{new}} = 0.5 - (-0.04)]$$

$$[w_{\text{new}} = 0.5 + 0.04 = 0.54]$$

Rounding to one decimal place:

$$[w_{\text{new}} = 0.5]$$

Surgical Precision Analysis

1. Impact of Update:

- The **0.04** adjustment ensures **smooth adaptation**, preventing abrupt movements.
- Maintains **stability** while optimizing force control.
- Further refinements over **10 iterations** will fine-tune the force application.

2. Safety Compliance:

- Any weight change **> 0.1** requires **human supervisor approval**.
- The **rounded value (0.5)** adheres to strict **safety certification limits**.

Final Answer:

0.5

Thus, the correct answer is:

✓ B) 0.5

29. CNN

Type: MCQ

Story: Tesla's Full Self-Driving v12 processes 224x224x3 camera frames through:

1. Conv Layer (64 filters, 3x3 kernel, stride=1, padding=1)
2. MaxPool (2x2 window, stride=2)

The chip designer needs to verify memory allocation matches the tensor shape after these layers. What configuration is correct?

- A) 112x112x64 (correct post-pooling output)
- B) 224x224x64 (convolution output before pooling)
- C) 112x112x3 (would discard learned features)
- D) 56x56x64 (over-aggressive downsampling)

Answer: A) 112x112x64

Autonomous Driving Impact:

1. **Conv Layer Math:** $(224 - 3 + 2 \times 1) / 1 + 1 = 224 \rightarrow$ preserves resolution
 2. **MaxPool Effect:** $\lfloor 224 / 2 \rfloor = 112 \rightarrow$ reduces computation by 75%
 3. **Why This Matters:**
 - 112px maintains enough detail for sign recognition
 - 64 channels capture critical edge/orientation features
 - Enables real-time processing at 60fps
-

30. SVM

Type: NAT

Story: Lockheed Martin's missile defense SVM classifies warheads (class 1) vs decoys (class - 1) with perfect separability. How many critical data points touch the decision boundary in the final model? (Integer)

Answer: 2

Military-Grade Classification:

1. Hard-Margin Dynamics:

- Exactly 1 warhead example defines "safe" boundary
- Exactly 1 decoy example defines "threat" boundary
- All other points are irrelevant to the margin

2. Operational Significance:

- Enables 99.99% accuracy with minimal storage
- Decision function depends only on these 2 vectors
- Survives even if 90% of training data is lost

Deployment Example:

- AEGIS system uses this property for shipboard defense
- 2 support vectors consume just 0.2KB memory
- Processes 100,000 classifications