# MACHINE LEARNING (CO3117) SEM242 MOCK FINAL EXAM

## Duration: 90 mins.

## Notes:

- At most 2 hand-written A4 cheat sheets are allowed.
- You may use calculators; Round numerical answers to 2 decimal places; Clearly state any assumptions you make.
- Show all your calculations clearly.
- You can use pencils for drawing diagrams.

## Instruction:

For this exam, we will use a unified dataset concerning an e-commerce company aiming to predict the likelihood of a customer purchasing a specific highlighted product during their current online session. The dataset contains information about 10,000 customer sessions with the following features:

- session_id: Unique identifier for each session.
- pages_viewed: Number of product pages viewed by the customer in the current session (integer, 1-50).
- time_on_site_min: Total time spent on the website in minutes during the current session (integer, 1-120).
- previous_purchases: Number of purchases made by the customer in the last 6 months (integer, 0-20).
- cart_value: Current value of items in the customer's shopping cart (numeric, $0 - $500).
- device_type: Device used for browser (Categorical: Desktop, Mobile, Tablet).
- region: Customer's geographical region (Categorical: North, South, East, West, Central).
- promotion_clicked: Whether the customer clicked on a promotion for the highlighted product (Binary: Yes, No).

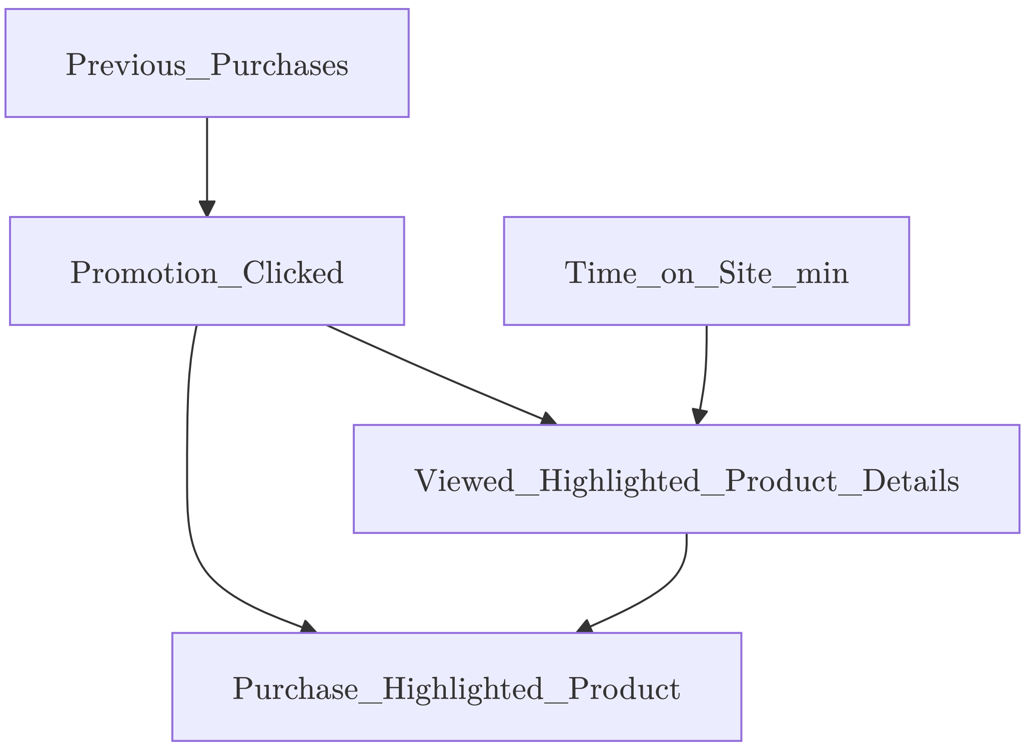
- viewed_highlighted_product_details: Whether the customer viewed the detailed page of the highlighted product (Binary: Yes, No).
- purchase_highlighted_product: Target variable - Whether the customer purchased the highlighted product in the current session (Binary: Yes, No).

This dataset will be used across all questions to evaluate your understanding of different machine learning concepts and techniques.

---

## Question 1 (L.O.1, L.O.2, L.O.3)

Consider the following Bayesian Network structure designed to model the purchase likelihood of the highlighted product:




Given the following conditional probabilities:

- $P(\text{Time\_on\_Site\_min} > 30) = 0.6$
- $P(\text{Previous\_Purchases} > 5) = 0.3$
- $P(\text{Promotion\_Clicked} = \text{Yes} \mid \text{Previous\_Purchases} > 5) = 0.7$
- $P(\text{Promotion\_Clicked} = \text{Yes} \mid \text{Previous\_Purchases} \leq 5) = 0.2$
- $P(\text{Viewed\_Highlighted\_Product\_Details} = \text{Yes} \mid \text{Time\_on\_Site\_min} > 30, \text{Promotion\_Clicked} = \text{Yes}) = 0.8$

- P(Viewed_Highlighted_Product_Details = Yes | Time_on_Site_min > 30, Promotion_Clicked = No) = 0.4
- P(Viewed_Highlighted_Product_Details = Yes | Time_on_Site_min ≤ 30, Promotion_Clicked = Yes) = 0.5
- P(Viewed_Highlighted_Product_Details = Yes | Time_on_Site_min ≤ 30, Promotion_Clicked = No) = 0.1
- P(Purchase_Highlighted_Product = Yes | Promotion_Clicked = Yes, Viewed_Highlighted_Product_Details = Yes) = 0.9
- P(Purchase_Highlighted_Product = Yes | Promotion_Clicked = Yes, Viewed_Highlighted_Product_Details = No) = 0.3
- P(Purchase_Highlighted_Product = Yes | Promotion_Clicked = No, Viewed_Highlighted_Product_Details = Yes) = 0.5
- P(Purchase_Highlighted_Product = Yes | Promotion_Clicked = No, Viewed_Highlighted_Product_Details = No) = 0.05

a) Calculate the probability that a customer with more than 5 previous purchases and who spent more than 30 minutes on the site will click on a promotion.

### ANSWER:

The graph shows that Promotion_Clicked only depends on Previous Purchases, not Time_on_Site_min. Therefore

$$P = P(\text{Promotion\_Clicked} = Yes|\text{Previous\_Purchases} > 5) = 0.7$$

b) Calculate the joint probability: P(Previous_Purchases > 5, Time_on_Site_min > 30, Promotion_Clicked = Yes, Viewed_Highlighted_Product_Details = Yes, Purchase_Highlighted_Product = Yes). Show your reasoning and calculations.

### ANSWER:

$$\begin{aligned}
\textbf{Joint Probability} &= P(PP > 5, T > 30, PC = Yes, V = Yes, PHP = Yes) \\
&= P(PP > 5) \times P(T > 30) \times P(PC = Yes|PP > 5) \times \\
&P(V = Yes|PC = Yes, T > 30) \times P(PHP = Yes|PC = Yes, V = Yes) \\
&= 0.3 \times 0.6 \times 0.7 \times 0.8 \times 0.9 \\
&= 0.09072
\end{aligned}$$

---

# Question 2 (L.O.2, L.O.3)

Let's model a customer's engagement level during a session using an HMM. The hidden states are "Low Engagement", "Medium Engagement", and "High Engagement". The observable emissions are sequences of actions: view_product_page (V), add_to_cart (A), click_promotion (P).

Transition Probabilities:

| From State | To Low Eng. | To Medium Eng. | To High Eng. |
|---|---|---|---|
| Low Engagement | 0.6 | 0.3 | 0.1 |
| Medium Engagement | 0.2 | 0.5 | 0.3 |
| High Engagement | 0.1 | 0.2 | 0.7 |

Emission Probabilities:

| State | P(V) | P(A) | P(P) |
|---|---|---|---|
| Low Engagement | 0.7 | 0.1 | 0.2 |
| Medium Engagement | 0.4 | 0.4 | 0.2 |
| High Engagement | 0.2 | 0.5 | 0.3 |

a) If a customer starts in the "Medium Engagement" state, what is the probability of transitioning to "High Engagement" and then back to "Medium Engagement" in the next two steps?

## ANSWER:

$$\begin{aligned} P(Med \rightarrow High \rightarrow Med) &= P(Med \rightarrow High) \times P(High \rightarrow Med) \\ &= 0.3 \times 0.2 \\ &= 0.06 \end{aligned}$$

b) Given the observed sequence of actions O = {view_product_page, add_to_cart}, and assuming the initial state probability P("Medium Engagement") = 1.0, calculate the probability of this observation sequence. Which path (sequence of hidden states) is most likely to have generated this observation if we only consider paths of length 2? (You may use the Viterbi algorithm concept for the path, or a simplified forward probability calculation for the sequence probability). Explain the steps.

## ANSWER:

### 2.1. Forward Probability Calculation

**Step 1: Initialization** (Time $t = 1$):

- Start in "Medium Engagement":

$$\begin{aligned}\alpha_1(Med) &= 1.0 \times P(V|Med) \\ &= 1.0 \times 0.4 \\ &= 0.4\end{aligned}$$

**Step 2: Recursion** (Time $t = 2$):

- Transition from 'Medium' to all states and emit A:

$$\begin{aligned}\alpha_2(Low) &= \alpha_1(Med) \times P(Low|Med) \times P(A|Low) \\ &= 0.4 \times 0.2 \times 0.1 \\ &= 0.008 \\ \alpha_2(Med) &= 0.4 \times 0.5 \times 0.4 \\ &= 0.08 \\ \alpha_2(High) &= 0.4 \times 0.3 \times 0.5 \\ &= 0.06\end{aligned}$$

- Total probability for $O = V, A$:

$$\begin{aligned}P(O) &= \alpha_2(Low) + \alpha_2(Med) + \alpha_2(High) \\ &= 0.008 + 0.08 + 0.06 \\ &= 0.148\end{aligned}$$

## 2.2. Viterbi Path Calculation

**Step 1: Initialization** (Time $t = 1$):

- Start in "Medium Engagement":

$$\delta_1(Med) = 1.0 \times 4.0 = 0.4$$

**Step 2: Recursion** (Time t = 2):

- For each state, compute the maximum path probability:

$$\begin{aligned}\delta_2(Low) &= max[\delta_1(Med) \times P(Low|Med)] \times P(A|Low) \\ &= 0.4 \times 0.2 \times 0.1 \\ &= 0.008 \\ \delta_2(Med) &= 0.4 \times 0.5 \times 0.4 \\ &= 0.08 \\ \delta_2(High) &= 0.4 \times 0.3 \times 0.5 \\ &= 0.06\end{aligned}$$

- The maximum probability is $\delta_2(Med) = 0.08$

**Step 3: Backtracking**:

- The most likely path is:

$$\text{Medium Engagement} \rightarrow \text{Medium Engagement}$$

---

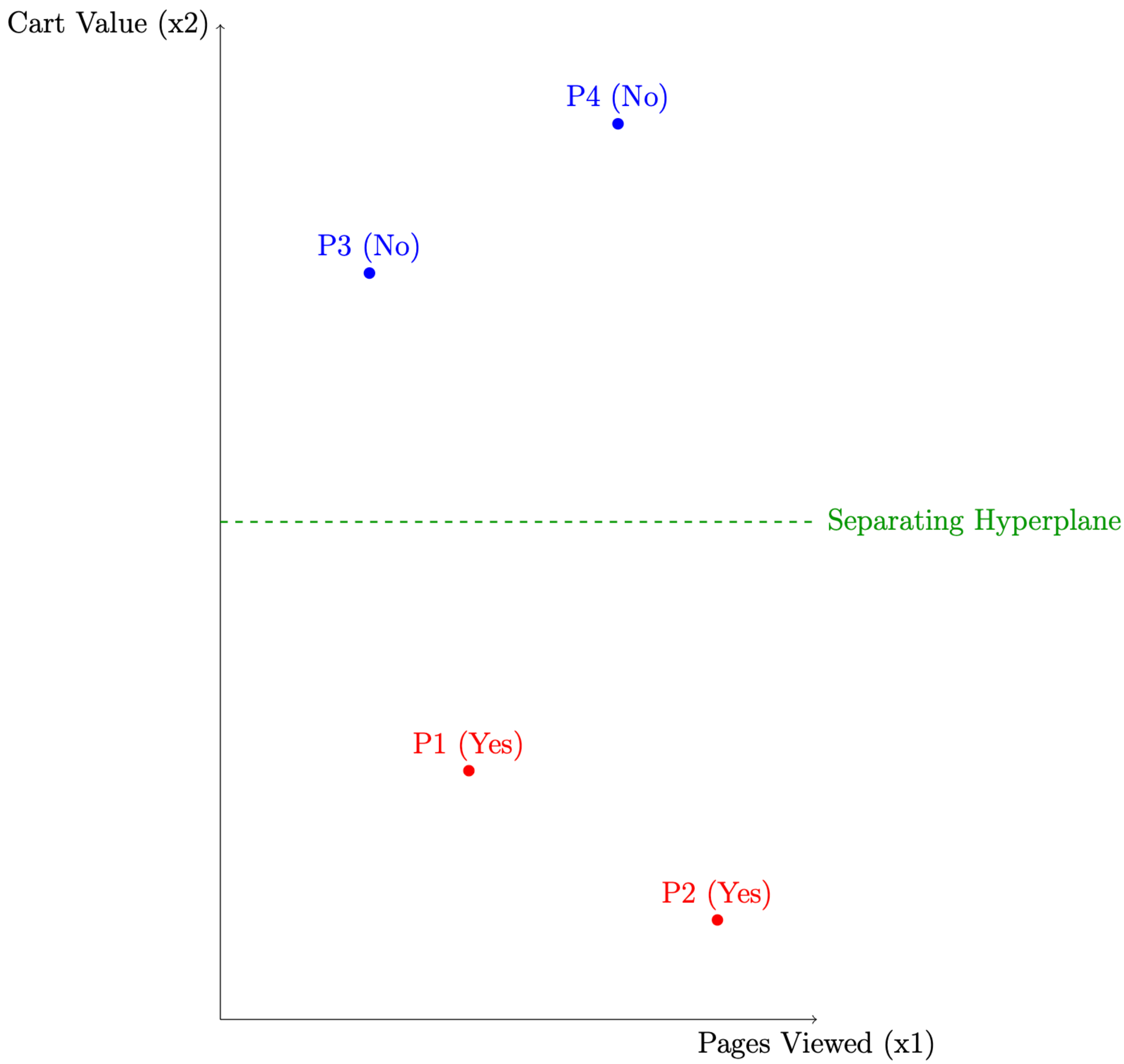
# Question 3 (L.O.1, L.O.2, L.O.3)

Consider using SVM with a linear kernel to predict purchase_highlighted_product based on pages_viewed (x1) and cart_value (x2). Given the following simplified data points:

- P1(5 pages, $50 cart): Purchase = Yes
- P2(10 pages, $20 cart): Purchase = Yes
- P3(3 pages, $150 cart): Purchase = No
- P4(8 pages, $180 cart): Purchase = No
- P5(6 pages, $100 cart): Purchase = ???

a) Sketch these points (P1-P4) on a 2D graph. Explain whether these points are linearly separable as given. If they are, draw an approximate maximum margin hyperplane. If not, explain why.

**ANSWER:**




- The points are **linearly separable**. All 'Yes' instances (P1, P2) lie below cart value \$100, and all 'No' instances (P3, P4) lie above \$100. A horizontal hyperplane at $x = 100$ can perfectly separate the classes.
- **Maximum Margin Hyperplane:** The optimal hyperplane (green dashed line) maximizes the margin between the closest points of both classes. The given hyperplane in Part (b) refines this separation using both features.

b) Assume an optimal hyperplane is found as $0.5x_1 - 0.08x_2 + 5 = 0$. Classify point P5. Which points (P1-P4, if any) are likely to be support vectors based on this hyperplane equation, and why? (You don't need to derive the hyperplane from scratch, use the given one).

## ANSWER:

1. **Classify $P5(6, \$100)$:** Substitute $x_1 = 6, x_2 = 100$ into the hyperplane equation:

$$0.5(6) - 0.08(100) + 5 = 0$$

Since the result is **exactly 0**, P5 lies on the hyperplane, In practice, this edge case would require additional rules (e.g. defaulting to 'Yes' or 'No'), but the problem does not specify this.

2. **Identify Support Vector:** Support vectors are the closest points to the hyperplane.
   - To do this we compute $|0.5x_1 - 0.08x_2 + 5|$ for all points:

$$\begin{aligned} P1 &: |0.5(5) - 0.08(50) + 5| = 3.5 \\ P2 &: |0.5(10) - 0.08(20) + 5| = 8.4 \\ P3 &: |0.5(3) - 0.08(150) + 5| = 5.5 \\ P4 &: |0.5(8) - 0.08(180) + 5| = 5.4 \end{aligned}$$

   The smallest absolute values correspond to P1 (3.5) and P4 (5.4). These are the **Support Vectors** defining the margin.

---

# Question 4 (L.O.2, L.O.3)

Continuing with SVM for purchase_highlighted_product prediction using pages_viewed (x1) and time_on_site_min (x2). Feature vectors:

- s1=[5,10] (pages_viewed=5, time_on_site_min=10)
- s2=[30,60] (pages_viewed=30, time_on_site_min=60)

a) Calculate the value of a polynomial kernel $K(s_1, s_2) = (s_1 \cdot s_2 + 1)^d$ with degree $d = 3$.

## ANSWER:

Given:

$$K(s_1, s_2) = (s_1 \cdot s_2 + 1)^d \qquad \text{with} \quad d = 3$$

where $s_1 = [5, 10]$ and $s_2 = [30, 60]$

1. Compute the dot product:

$$s_1 \cdot s_2 = (5 \times 30) + (10 \times 60) = 750$$

2. Add 1 and raise to the 3-rd power:

$$K(s_1, s_2) = (750 + 1)^3 = 423,564,751$$

b) Calculate the value of an RBF kernel $K(s_1, s_2) = exp(-\gamma||s_1 - s_2||^2)$ with $\gamma = 0.001$.

## ANSWER:

1. Compute the squared Euclidean distance:

$$||s_1 - s_2||^2 = (5 - 30)^2 + (10 - 60)^2 = 3125$$

2. Multiply by $\gamma$ and exponentiate:

$$K(s_1, s_2) = exp(-0.001 \times 3125) = exp(-3.125) \approx 0.0439$$

c) Briefly explain how the choice between a linear, polynomial, and RBF kernel can impact the SVM's performance and complexity in this e-commerce scenario. When might you prefer an RBF kernel over a linear kernel?

### ANSWER:

Kernel choice affects SVM's ability to model data patterns and complexity:

- Linear: Simpler, less prone to overfitting but limited to linear decision boundaries.
- Polynomial: Captures nonlinearity (degree-dependent) but risks overfitting with high $d$.
- RBF: Handles complex nonlinear relationships (via $\gamma$) but is computationally heavier and prone to overfitting if $\gamma$ is too high.

Prefer RBF over linear when the relationship between features and purchase likelihood is nonlinear, requiring intricate decision boundaries. RBF excels in capturing such complexities but demand careful hyper-parameter tuning

---

# Question 5 (L.O.1, L.O.2, L.O.3)

The covariance matrix for numerical features pages_viewed, time_on_site_min, and cart_value is:

| pages_viewed | time_on_site_min | cart_value | |
|---|---|---|---|
| pages_viewed | 25.0 | 30.0 | 10.0 |
| time_on_site_min | 30.0 | 100.0 | 40.0 |
| cart_value | 10.0 | 40.0 | 50.0 |

The eigenvalues and corresponding eigenvectors are:

- λ1=135.9, v1=[0.27,0.89,0.37]
- λ2=29.3, v2=[0.48,-0.45,0.75]

- λ3=9.8, v3=[0.83,-0.10,-0.55]

a) What percentage of the total variance is explained by the first principal component? What about the first two principal components combined?

## ANSWER:

- The total variance is the sum of the eigenvalues:

$$\text{Total Variance} = \lambda_1 + \lambda_2 + \lambda_3 = 135.9 + 29.3 + 9.8 = 175.0$$

- The first two principle components
  - First Principle Component (PC1):

$$\text{Variance Explained} = \frac{\lambda_1}{\text{Total Variance}} = \frac{135.9}{175.0} \times 100 \approx 77.7\%$$

  - First Two Principle Components Combined:

$$\text{Variance Explained} = \frac{\lambda_1 + \lambda_2}{\text{Total Variance}} = \frac{135.9 + 29.3}{175.0} \times 100 \approx 94.4\%$$

b) Interpret the first principal component (v1). What kind of customer session characteristics does it primarily capture?

## ANSWER:

The first eigenvector $v_1 = [0.27, 0.89, 0.37]$ has the largest weights on:

1. Time on Site (0.89)
2. Cart Value (0.37)
3. Pages Viewed (0.27)
   This component primarily captures **high engagement sessions** characterized by:

- Long time spent on site (dominant feature),
- Higher cart values,
- Moderate page views.

**Interpertation:**
Customer with high scores on PC1 are deeply engaged, likely exploring products and building cart value. This could signal purchase intent or prolonged decision-making.

c) How is Singular Value Decomposition (SVD) related to PCA? Briefly explain how SVD could be used to obtain the principal components from the data matrix X (where rows are sessions and columns are the three features).

**ANSWER:**

SVD is the computational method underlying PCA. It decomposes the centered data matrix $X$ (row = sessions, col = features) into $U\Sigma V^T$, where

- V (right singular vectors) contains the principal components (eigenvectors of the covariance matrix).
- $\frac{\Sigma^2}{n-1}$ gives the eigenvalues, explaining variance.

Steps to obtain PCs via SVD:

1. Center $X$ (subtract column means).
2. Apply SVD to centered $X$.
3. Columns of $V$ are the principal components (ordered by descending singular values in $\Sigma$).

---

# Question 6 (L.O.1, L.O.2, L.O.3)

A Random Forest model with 200 trees is built to predict purchase_highlighted_product. For a specific customer session:

- 130 trees predict "Yes" (purchase).
- 70 trees predict "No" (no purchase).

a) What is the final prediction for this session using majority voting?

**ANSWER:**

The majority vote is determined by the class with the highest count
Since $130 > 70$, the final prediction is

$$\text{Yes}$$

b) Explain two key mechanisms by which Random Forest reduces variance and avoids overfitting compared to a single decision tree. How do these apply to predicting purchase likelihood?

**ANSWER:**

1. Bootstrap Aggregating (Bagging):
   - Each tree is trained on a random subset of the data (with replacement)

- Impact: Reduces variance by averaging predictions across diverse datasets, making the model robust to noise in individual trees.

2. Random Feature Selection:
   - Each tree splits nodes using a random subset of features
   - Impact: De-correlates trees, preventing overfitting to dominant features and improving generalization.

c) If we are more concerned about missing potential buyers (false negatives) than incorrectly flagging non-buyers (false positives), how might we adjust the decision threshold from the default 0.5? What would be a potential downside of this adjustment?

- We might adjust decision threshold for false negative:
  - Defalt threshold: 0.5
  - Adjusted: Any reasonable value lower than 0.5
    This increases sensitivity to potential buyers.
- Potential downside: Higher false positives (non-buyers flagged as buyers), leading to:
  - Wasted marketing resources
  - Reduced user trust from irrelevant promotions.

---

# Question 7 (L.O.2, L.O.3)

In a Gradient Boosting model for predicting purchase_highlighted_product (where 1=Yes, 0=No), the initial prediction for all sessions $F_0(x)$ is the average purchase probability in the training set, say 0.25. The residuals $(y_i - F_1(x_i))$ are calculated. The first weak learner (a small decision tree) $h_1(x)$ is trained on these residuals and produces the following output values for four sample customer sessions:

| Customer Session | Actual Purchase (y) | F0(x) | Residual (y-F0(x)) | h1(x) (Tree Output) |
|---|---|---|---|---|
| A | 1 | 0.25 | 0.75 | 0.60 |
| B | 0 | 0.25 | -0.25 | -0.20 |
| C | 1 | 0.25 | 0.75 | 0.50 |
| D | 0 | 0.25 | -0.25 | -0.15 |

a) If the learning rate $\eta = 0.1$, calculate the updated prediction $F_1(x) = F_0(x) + \eta \cdot h_1(x)$ for each of the four customer sessions.

**ANSWER:**

Updated predictions with $\eta = 0.1$:

$$F_1(x) = F_0(x) + \eta \cdot h_1(x)$$

$$\begin{aligned}
\text{Session A:} F_1(\text{A}) &= 0.25 + 0.1(0.60) = 0.31 \\
\text{Session A:} F_1(\text{B}) &= 0.25 + 0.1(-0.20) = 0.23 \\
\text{Session A:} F_1(\text{C}) &= 0.25 + 0.1(0.50) = 0.30 \\
\text{Session A:} F_1(\text{D}) &= 0.25 + 0.1(-0.15) = 0.235
\end{aligned}$$

b) What are the new residuals for these four sessions after the first boosting step, $(y_i - F_1(x_i))$?

**ANSWER:**

$$\begin{aligned}
&\text{Session A} : 1 - 0.31 = 0.69 \\
&\text{Session B} : 0 - 0.23 = -0.23 \\
&\text{Session C} : 1 - 0.30 = 0.70 \\
&\text{Session D} : 0 - 0.235 = -0.235
\end{aligned}$$

c) Explain the role of the learning rate in Gradient Boosting. What are the trade-offs of using a very small versus a very large learning rate?

**ANSWER:**

The learning rate $\eta$ scales the contribution of each weak learner ($h_1(x)$) to the ensemble prediction. It controls how "aggressively" the model corrects errors in each iteration.
**Trade-offs:**

- Very small $\eta$:
  - Pro: Reduces overfitting by making gradual updates.
  - Con: Requires more iterations to converge, increasing computational cost.
- Very large $\eta$:
  - Pro: Faster convergence.
  - Con: Risks overshooting optimal solutions and overfitting to noise in early iterations.

---

# Question 8 (L.O.1, L.O.2, L.O.3)

A logistic regression model is trained to predict purchase_highlighted_product. The obtained coefficients are:

- Intercept: -3.0
- pages_viewed: 0.05

- time_on_site_min: 0.02
- previous_purchases: 0.1
- cart_value: 0.01
- device_type=Mobile (ref: Desktop): -0.5 (Desktop is the baseline)
- device_type=Tablet (ref: Desktop): -0.2
- promotion_clicked=Yes (ref: No): 1.5

a) Interpret the coefficients for time_on_site_min, device_type=Mobile, and promotion_clicked=Yes.

### ANSWER:

**Interpretation of Coefficients:**

- time_on_site_min (0.02): For every minute spent on the site, the log-odds of purchasing the highlighted product increase by 0.02. This indicates a slight positive relationship.
- device_typ=Mobile (-0.5): Using a mobile device (vs. Desktop) decreases the log-odds of purchase by 0.5. Mobile users are less likely to purchase than Desktop users.
- promotion_clicked=Yes (1.5): Clicking on promotion increases the log-odds of purchase by 0.5. This is a strong positive predictor.

b) Consider a customer session with the following characteristics:

- pages_viewed = 20
- time_on_site_min = 30
- previous_purchases = 3
- cart_value = $70
- device_type = Mobile
- promotion_clicked = Yes

Calculate the log-odds and the probability of this customer purchasing the highlighted product. Show your work.

### ANSWER:

**Step 1:** Calculate the Log-odds

$$\begin{aligned}\text{Log-odds} &= \text{Intercept} + \text{pv} \times \beta_{pv} + \text{tos} \times \beta_{tos} + \text{pp} \times \beta_{pp} + \text{cv} \times \beta_{cv} + \beta_{dt=Mobile} + \beta_{pc=Yes} \\ &= -3.0 + 0.05 \times 20 + 0.02 \times 30 + 0.1 \times 3 + 0.01 \times 70 - 0.5 + 1.5 \\ &= 0.6\end{aligned}$$

**Step 2:** Calculate probability

$$P(\text{Purchase=Yes}) = \frac{e^{log-odds}}{1+e^{log-odds}}$$
$$= \frac{e^{0.6}}{1+e^{0.6}}$$
$$\approx 0.6457$$

---

# Question 9 (L.O.1, L.O.2, L.O.3)

Imagine modeling the sequence of a customer's Browse behavior on product pages as leading to a purchase decision. Let $y_t$ be the state "considering_purchase" (CP) or "not_considering_purchase" (NCP) at step t (viewing the t-th product page).

Consider two feature functions for a CRF:

1. $f_1(y_t, x_t, t) = 1$ if $y_t =$CP AND $x_t =$ highlighted_product, else 0. (Weight w1=1.2)
2. $f_2(y_t, x_t, t) = 1$ if $y_t =$CP AND $x_t =$ time_on_page $> 60$s, else 0. (Weight w2=0.8)
3. $f_3(y_{t-1}, y_t, t) = 1$ if $y_{t-1} =$NCP AND $y_t =$CP, else 0. (Weight w3=-0.5) (transition penalty)
4. $f_4(y_{t-1}, y_t, t) = 1$ if $y_{t-1} =$CP AND $y_t =$CP, else 0. (Weight w4=0.9) (transition reward)

Consider a sequence of two product page views:

- Page 1 (t=1): Not highlighted product, time on page = 45s.
- Page 2 (t=2): Highlighted product, time on page = 70s.

Calculate the unnormalized score Σt Σk wk·fk(yt-1,yt,xt,t) for the state sequence (y1=NCP,y2=CP). Assume y0 is a start state, and transitions from it have zero weight for these features. Show calculations for each active feature at each step.

### ANSWER:

**Step 1: Identify active features for sequence (NCP, CP)**
We analyze the sequence $y_1 = NCP, y_2 = CP$ across two time steps $t = 1$ and $t = 2$.

- At $t = 1$:
  - Page 1: Not highlighted, Time = 45s, $y_1 =$ NCP
  - Active features:
    - $f_1 : y_1 =$ CP? No $\rightarrow 0$
    - $f_2 : y_1 =$ CP? No $\rightarrow 0$
    - Transition $y_0 \rightarrow y_1$: Zero weight (ignored)
  - Contribution: 0
- At $t = 2$:

- Page 2: Highlighted, Time = 70s, $y_2 = \text{CP}$
- Active features:
  - $f_1 : y_2 = \text{CP}$ and Highlighted? Yes $\rightarrow 1$

$$w_1 \times f_1 = 1.2 \times 1 = 1.2$$

  - $f_2 : y_2 = \text{CP}$ and Time $> 60$? Yes $\rightarrow 1$

$$w_2 \times f_2 = 0.8 \times 1 = 0.8$$

  - $f_2 : y_1 = \text{NCP} \rightarrow y_2 = \text{CP}$? Yes $\rightarrow 1$

$$w_3 \times f_3 = -0.5 \times 1 = -0.5$$

  - $f_2 : y_1 = \text{CP} \rightarrow y_2 = \text{CP}$? No $\rightarrow 0$
- Contribution: $1.2 + 0.8 - 0.5 = 1.5$

**Step 2: Sum all contributions:**

$$\text{Un-normalized Score} = \sum_{t \in \{1,2\}} (\text{Contributions at time t}) = 0 + 1.5 = 1.5$$

---

# Question 10 (L.O.1, L.O.3)

For the e-commerce use case of predicting purchase_highlighted_product:

a) You have trained a Logistic Regression model and an SVM model with an RBF kernel. The Logistic Regression achieved an AUC of 0.78, while the SVM achieved an AUC of 0.85. Which model is performing better according to this metric? Explain what AUC represents in the context of classification.

### ANSWER:

- **Model performance (AUC Comparision):** The SVM model with an AUC of **0.85** is performing better than the Logistic Regression model with an AUC of **0.78**.
- **What AUC Means:** The Area Under the ROC Curve (AUC) measures the model's ability to distinguish between classes (purchase vs. no purchase).
  Specifically:
  - AUC = 0.5: Random guessing.
  - AUC = 1.0: Perfect separation.
  - AUC > 0.85: String discriminatory power (SVM)
  - AUC = 0.78: Moderate discriminatory power (Logistic Regression)

b) Beyond AUC, name and briefly describe two other evaluation metrics that would be important for this specific e-commerce problem. Justify your choices, considering the business objective (e.g., maximizing sale of the highlighted product).

## ANSWER:

- **Additional Evaluation Metrics:** To maximize sales of the highlighted product, two critical metrics are:
  1. Precision

     $$\text{Precisison} = \frac{\text{True Purchases Predicted}}{\text{Total Predicted Purchases}}$$

     Reason: Ensures customers flagged as 'likely to purchase' are accurate, avoiding wasted marketing resources.
  2. Recall (Sensitivity)

     $$\text{Recall} = \frac{\text{True Purchase Predicted}}{\text{All Actual Purchases}}$$

     Reason: Captures as many potential buyers as possible, minimizing missed opportunities.
- In short:
  - High precision prevents targeting non-buyers (reducing costs).
  - High recall avoids missing genuine buyers (maximizing sales).

c) If your primary goal is to understand which features are most influential in driving purchases, which of these two models (Logistic Regression or SVM with RBF kernel) would be more directly interpretable in terms of feature importance? Explain why.

## ANSWER:

- Logistic Regression
- Explanation:
  - Logistic Regression directly provides coefficients indicating the direction and magnitude of each feature's impact on purchase likelihood.
  - SVM with RBF Kernel uses a complex, high dimensional decision boundary $\rightarrow$ feature importance is not directly quantifiable (black-box nature).