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:

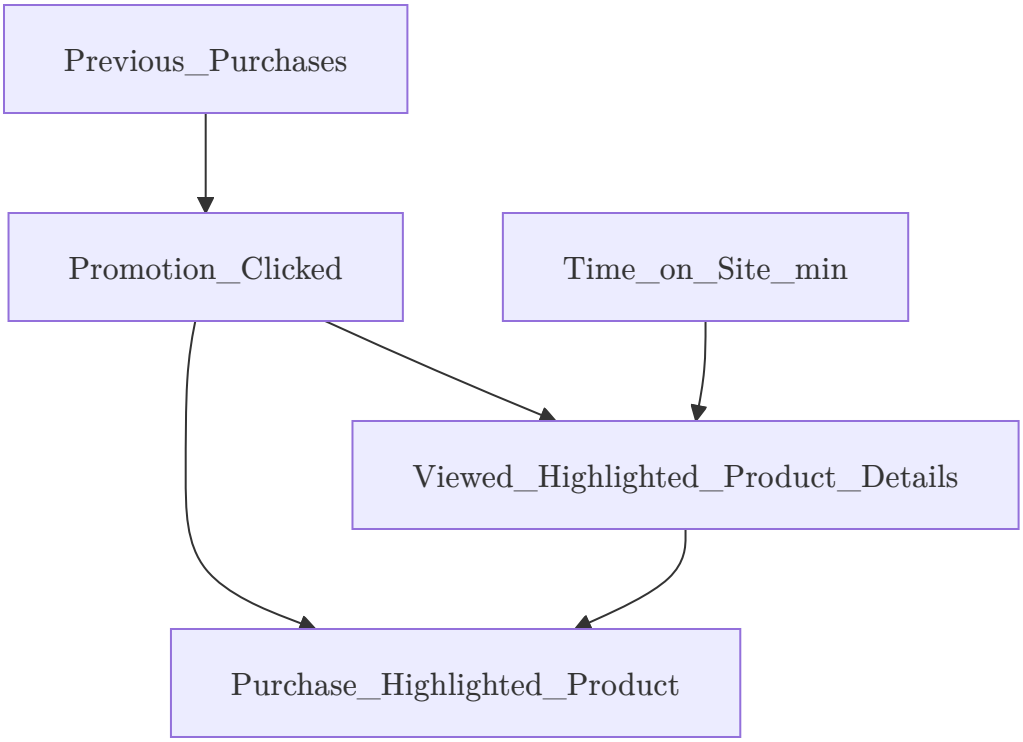
- session_id: Unique identifier for each session.
- pages_viewed: Number of product pages viewed by the customer in the current session (integer, 1-50).
- time_on_site_min: Total time spent on the website in minutes during the current session (integer, 1-120).
- previous_purchases: Number of purchases 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:

- $P(\text{Time\_on\_Site\_min} > 30) = 0.6$
- $P(\text{Previous\_Purchases} > 5) = 0.3$
- $P(\text{Promotion\_Clicked} = \text{Yes} \mid \text{Previous\_Purchases} > 5) = 0.7$
- $P(\text{Promotion\_Clicked} = \text{Yes} \mid \text{Previous\_Purchases} \leq 5) = 0.2$
- $P(\text{Viewed\_Highlighted\_Product\_Details} = \text{Yes} \mid \text{Time\_on\_Site\_min} > 30, \text{Promotion\_Clicked} = \text{Yes}) = 0.8$
- $P(\text{Viewed\_Highlighted\_Product\_Details} = \text{Yes} \mid \text{Time\_on\_Site\_min} > 30, \text{Promotion\_Clicked} = \text{No}) = 0.4$

- 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.

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.

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

State	P(V)	P(A)	P(P)
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?

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.

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.

b) Assume an optimal hyperplane is found as $0.5x1-0.08x2+5=0$. 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).

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 $K(s1,s2)=(s1 \cdot s2+1)d$ with degree $d=3$.

b) Calculate the value of an RBF kernel $K(s1,s2)=\exp(-\gamma||s1-s2||2)$ with $\gamma=0.001$.

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?

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:

- $\lambda1=135.9$, $v1=[0.27,0.89,0.37]$
- $\lambda2=29.3$, $v2=[0.48,-0.45,0.75]$
- $\lambda3=9.8$, $v3=[0.83,-0.10,-0.55]$

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

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

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).

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?

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?

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?

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 F0(x) is the average purchase probability in the training set, say 0.25. The residuals (yi-F0(xi)) are calculated. The first weak learner (a small decision tree) h1(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 η=0.1, calculate the updated prediction F1(x)=F0(x)+η·h1(x) for each of the four customer sessions.

b) What are the new residuals for these four sessions after the first boosting step, i.e., (yi-F1(xi))?

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?

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.

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.

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 yt 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. f1(yt,xt,t)=1 if yt=CP AND x_t = text(highlighted_product), else 0. (Weight w1=1.2)
2. f2(yt,xt,t)=1 if yt=CP AND x_t = text(time_on_page) > 60s, else 0. (Weight w2=0.8)
3. f3(yt-1,yt,t)=1 if yt-1=NCP AND yt=CP, else 0. (Weight w3=-0.5) (transition penalty)
4. f4(yt-1,yt,t)=1 if yt-1=CP AND yt=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.

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.

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).

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.