Mock Exam Questions - Module 3

Short Answer

1. Prove that the function $k : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ defined by $k(u, v) = (2 + u \cdot v)^2$ is a valid kernel.

2. When building a classification model, how does a single decision tree compare to a random forest ensemble in terms of interpretability, computational cost, and susceptibility to overfitting?

3. After training a decision tree to predict whether a bank transaction is fraudulent, how would you classify a brand-new transaction?

4. In a manufacturing plant, let event A be that a randomly selected component is produced by Machine X, with $P(A) = 0.6$. Let event B be that the component passes the final quality inspection. Given that $P(B|A) = 0.3$ and $P(A|B) = 0.45$, determine the overall probability $P(B)$ that a randomly selected component passes inspection.

5. In an email spam detection system, explain why using an ensemble of decision trees trained on random subsets of the emails and features typically achieves higher classification accuracy than relying on a single decision tree.

6. In a medical study classifying patients into disease categories using the CART algorithm, what are the main limitations or drawbacks of applying CART in this context?

Medium Answer

1. In the context of building a decision tree to classify emails into "spam," "promotional," and "important," define the Gini Index and describe how it is computed and applied at each node to choose the attribute that yields the purest subsets.

2. When constructing a decision tree to classify emails as spam or not spam, how does the CART algorithm choose the optimal split at each node for classification tasks?

3. In an SVM training problem, the optimal dual variable vector is found to be $\alpha^* = (0, 0, 1.5, 0, 2.7, 0, 0, 3.1)^T$ for eight training examples indexed 1 through 8. Which of these examples correspond to the support vectors?

4. In building a decision tree to classify emails as spam or not based on word frequencies, define information gain and explain how it is used to choose the best feature to split on.

5. In a medical diagnosis application using a Naïve Bayes model, what fundamental assumption is made about the relationship among the various symptoms given a specific disease?

6. Starting from Bayes' theorem, derive the decision rule for a classifier that assumes conditional independence among features.

Long Answer

1. Describe in detail the key steps you would follow to implement a Naive Bayes classifier for detecting spam emails, from feature extraction through probability estimation to making the final classification.

2. In support vector machines, what is a kernel function and how does the kernel trick enable the model to perform non-linear classification by computing inner products in a transformed feature space without explicitly mapping the data?

3. Show that the function $K : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ defined by $K(x, y) = (2 + x \cdot y)^3$ is a valid kernel.

4. Consider the table below which contains data about second-hand smartphone listings. We want to build a decision tree classifier to predict whether a listing is a good deal to buy. We consider the two classes 'Accept' and 'Reject'. Build the decision tree using the ID3 algorithm and information gain.

No.	Price	Battery	Storage	Warranty	Deal
1	High	Excellent	64GB	Yes	Accept
2	Low	Good	64GB	No	Accept
3	Medium	Excellent	128GB	Yes	Accept
4	Low	Good	128GB	Yes	Reject
5	High	Poor	256GB	No	Reject
6	High	Excellent	256GB	No	Reject
7	Medium	Good	64GB	No	Reject

5. Use a naive Bayes classifier learned from this training set to predict the classification for these two test fruits:

   (a) A fruit that is Soft (+) and has Sweet aroma (+) (other features unknown).

   (b) A fruit that is Red-colored (+) and Heavy (+) (other features unknown).