

- 1 You are developing a machine learning system for a well known supermarket chain. You have collected a large amount of data on shoppers who use their loyalty card. Which of the following attributes are best modelled as categorical attributes

**Select one alternative:**

- ☐ If the customer owns a Volvo or not
- ☐ The age of the customer
- ☐ The BMI of the customer
- ☐ How much the customer spends per month.
- ☐ How much candy/sweets/fredagsgodis the customer buys

---

Maximum marks: 1

- 2 You are still working for the well known supermarket chain. You want to predict how much candy/sweets/fredagsgodis the customer will buy. Your training set includes all of the above information. Which of the following types of machine learning algorithm is most appropriate

**Select one alternative:**

- ☐ Unsupervised Learning
- ☐ Supervised Learning

---

Maximum marks: 1

- 3 You want to do a promotion with Volvo, and so you want to be able to predicate if a new customer owns a Volvo or not. This is an example of a

**Select one alternative:**

- ☐ Classification problem
- ☐ Regression problem

---

Maximum marks: 1

4 One of these algorithms is not suitable for classification.

**Mark which one it is**

- ☐ Support Vector Machines
- ☐ Linear Regression
- ☐ Logistic Regression
- ☐ Naive Bayes

---

Maximum marks: 1

5 You want to divide your customers into groups of similar types of customers so that you can do target marketing. Which of the following machine learning algorithms is most appropriate.

**Select one alternative:**

- ☐ Linear Regression
- ☐ Support Vector Machines
- ☐ K-means clustering
- ☐ Logistic Regression

---

Maximum marks: 1

6 Which of the following best describes a hyper-parameter?

**Select one alternative:**

- ☐ A setting in the machine learning algorithm that tells you what format the training data is in.
- ☐ Parameters of your learning algorithm that are not estimated from the training set.
- ☐ The size of the training data set.
- ☐ The number of false positives your algorithm gives on the training set.

---

Maximum marks: 1

7 The loss or error function is used to train a learning algorithm. In logistic regression is the loss function calculates the number of true positives on the test or validation set:

**Select one alternative:**

- ☐ True
- ☐ False

---

Maximum marks: 1

8 Naive Bayesian can only be used as a classification algorithm

**Select one alternative:**

- ☐ False
- ☐ True

---

Maximum marks: 1

**9** When would you use an algorithm such as principal component analysis (PCA) to reduce the dimension of your data:

**Select one alternative:**

- ☐ When your computer cannot load the data quick enough
- ☐ When you have too many training samples
- ☐ When there are a large number of features/attributes that are linear dependent on each other.
- ☐ When your data has a large number of features/attributes that are related by a non-linear transformations.

---

Maximum marks: 1

**10** Which of the following best describes over fitting.

**Select one alternative:**

- ☐ The value of the regularisation parameter in logistic regression.
- ☐ When the model learns features of the training set that do not generalise well to the test/validation set.
- ☐ The resulting value error/loss function on the training set after training is too high.
- ☐ When the learning algorithm takes too long to run

---

Maximum marks: 1

11

Using the bag of words model for spam. It would be possible train a logistic regression classifier to predict if a message is spam or not. (It does not matter how good a classifier it is, only if it is possible to use logistic regression or not.)

**Select one alternative**

- ☐ False
- ☐ True

---

Maximum marks: 1

12 During training the loss/error function always converges to 0.

**Select one alternative:**

- ☐ False
- ☐ True

---

Maximum marks: 1

13 K-fold cross validation is used for one of the following purposes

**Select one alternative:**

- ☐ To avoid over fitting by reducing the dependency on one training set.
- ☐ To guarantee that the loss/error function converges to 0.
- ☐ To train a classifier to decide if a lamb is sick or not.
- ☐ To make the learning algorithm faster by using smaller training sets.

---

Maximum marks: 1

**14** One of the following statements is true about K-means clustering learning algorithm.

**Select one alternative:**

- ☐ K-means clustering is a modified form a K-nearest neighbours.
- ☐ K-means clustering always converges to a global minimum
- ☐ It is often necessary to restart the K-means learning algorithm to avoid local minima .
- ☐ During training the number of clusters are gradually reduced until the optimal number of cluster is found.

---

Maximum marks: 1

**15** The ID3 learning algorithm for decision trees always learns an optimal decision tree

**Select one alternative:**

- ☐ False
- ☐ True

---

Maximum marks: 1

**16** What best describes the purpose of random forest algorithm for decision trees.

**Select one alternative:**

- ☐ After learning a random forest the trees are combined into one large tree that is guaranteed to be optimal.
- ☐ The random forest is a more efficient algorithm to do K-fold cross validation. One decision tree is learnt for each selected training set.
- ☐ In order to use a decision tree for regression randoms values are taken and a tree is trained to solve a classification problem for that value.
- ☐ Random forests overcome over fitting by learning lots of small trees that are trained on a subset of the features/attributes. The resulting decision is an ensemble based decision.

---

Maximum marks: 1

- 17** The features/attributes Money, Honey, Learning, Exam are 1 if a message contains the word and 0 otherwise. Repeats are not taken into account. Consider the following training set. The last column classifies if the message is spam or not.

Money	Honey	Learning	Exam	Spam YES/NO
1	1	0	0	YES
1	0	0	1	YES
0	1	0	0	YES
0	0	0	1	NO
0	0	1	1	NO

Given a message containing the words Exam and Honey calculate using Bayes' theorem the probability that the message is spam or not. You must show all your workings, and not just the answer. (2 points) Please do the best that you can with ASCII mathematics. I can read LaTeX source. The box expands to accommodate your answer.

**Fill in your answer here**

You receive a message containing the words Honey and Learning, again calculate using Bayes' theorem the probability that the message is spam or not. You must show all your workings, and not just the answer. Comment on why you might want to use Laplacian smoothing when calculating the probability of a message being spam or not. (3 points)

**Fill in your answer here**

---

Maximum marks: 5



- 18** Using the same data set as before (repeated here) you are to start the calculation of a decision tree. Again you are trying to classify if a message is Spam or not .

Money	Honey	Learning	Exam	Spam YES/NO
1	1	0	0	YES
1	0	0	1	YES
0	1	0	0	YES
0	0	0	1	NO
0	0	1	1	NO

First calculate the entropy of the data set. You must show all your workings. (2 points)

**Fill in your answer here**

For each feature calculate the information gain of that feature. You must show all your workings. (2 points)

**Fill in your answer here**

Which node will be at the top of your decision tree and why? (1 point)

**Fill in your answer here**

---

Maximum marks: 5

**19** What is one-hot encoding? Give an example and explain why it is necessary? (3 points)

**Fill in your answer here**

You are building a multi-class classifier using support vector machines. Your data set have five classes A,B,C,D,E describe two strategies for combining support vector machines to produce a classifier for the five classes. (2 points)

**Fill in your answer here**

Describe two strategies that you can use when you data set contains missing values for some attributes. (2 points)

**Fill in your answer here**

You are going to build a recommender system that predicts what film/movie a user is most likely to watch based on their past history. You are free to use supervised or unsupervised learning. Describe what sort of approach you would use. In particular describe what sort of data you would collect and what attributes you would use. (5 points)

**Fill in your answer here**

---

Maximum marks: 12