ⓘ 2021-Aug Instructions

This exam has two parts. A multiple choice section and some more general questions. Unlike some previous exams you should answer all parts of the exam (even if you are only aiming for a 3).

17 Marks are required for a.3.
22 Marks are required for a 4.
28 Marks are required for a 5.
I will be online 8:00 - 9:00 and 12:00-13:00 to answer questions.

Each multiple choice question is worth 1 point.

I will be online to answer clarification questions 8:00 - 9:00 and 12:00-13:00. You are advised to read the whole exam before you start answering the questions.

This is an individual closed book exam, and you are expected to treat it as one. Collaboration or using resources such as the internet or any other material is considered cheating.

1 2021-ML-June-Q3

You have a data-set labelled into two classes "True" and "False". Which of the following machine learning algorithms could you try without any modifications?
Select one alternative:

○ Logistic regression

○ K-means clustering.

○ Principle Component Analysis

○ Linear Regression

Maximum marks: 1

2 **2021-ML-June-Q4**

What probabilistic assumption do you make for naive Bayes to be correct?
Select one alternative:

○ Some feature are independent of each other.

○ All features are independent of each other.

○ No features are independent of each other.

○ Most features are independent of each other.

Maximum marks: 1

3 **2021-ML-Juine-Q2**

Which of these is *not* a technique to avoid overfitting?
Select one alternative:

○ Using a subset of features of your dataset.

○ Using a regularisation term in cost/error/loss function.

○ Using a machine learning algorithm with a simpler set of possible learned hypotheses.

○ K-means clustering

Maximum marks: 1

4 2021-ML-June-Q1

When you consider linear regression as a machine learning algorithm on two dimensional data implemented as gradient descent then which of the following statements is *true*.
Select one alternative:

○ The solution will always find a straight line that goes through all the training data points.

○ The straight line produced will minimise the mean squared error for each training data point.

○ If you have too many training data points then your algorithm will always overfit.

○ The straight line produced will have a gradient that is the average of all the data points.

Maximum marks: 1

5 2021-ML-June-Q6

One of these algorithms is *not* suitable for classification. Mark which one it is
Select one alternative:

○ Linear Regression

○ Logistic Regression

○ Support Vector Machines

○ Naive Bayes

Maximum marks: 1

6 2021-ML-June-Q5

Which of the following problems is *not* a regression problem?

Select one alternative

○ Predicting the probability that a message is spam or not.

○ Predicting the average temperature of mid summer day based on average temperatures of all days leading up to mid summer day for that year.

○ Given a students performance during the course, predicting if the student will pass or fail.

○ Predicting the final sale price of a house

Maximum marks: 1

7 2021-ML-June-Q8

Which of the following best describes a hyper-parameter?
Select one alternative:

○ The number of false positives your algorithm gives on the training set.

○ The size of the training data set.

○ 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.

Maximum marks: 1

8 **2021-ML-June-Q7**

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

Select one alternative

○ Logistic Regression

○ Support Vector Machines

○ K-means clustering

○ Linear Regression

Maximum marks: 1

9 **2021-ML-June-Q9**

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:

○ False

○ True

Maximum marks: 1

10 **2021-ML-June-Q10**

Logistic regression requires all variables to be categorica
Select one alternative:

○ True

○ False

Maximum marks: 1

11 **2021-ML-June-Q12**

Which of the following best describes over fitting.

Select one alternative:

○ The value of the regularisation parameter in logistic regression

○ The resulting value error/loss function on the training set after training is too high.

○ When the learning algorithm takes too long to run

○ When the model learns features of the training set that do not generalise well to the test/validation set.

Maximum marks: 1

12 2021-ML-June-Q11

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 your data has a large number of features/attributes that are related by a non-linear transformations.

○ When there are a large number of features/attributes that are linear dependent on each other.

○ When you have too many training samples.

Maximum marks: 1

13 2021-ML-June-Q14

During training a loss/error function *always* converges to 0.
Select one alternative:

○ True

○ False

Maximum marks: 1

14 2021-ML-June-Q13

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:

○ True

○ False

Maximum marks: 1

15 2021-ML-June-Q15

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

Select one alternative:

○ It is often necessary to restart the K-means learning algorithm to avoid local minima

○ K-means clustering is a modified form a K-nearest neighbours.

○ K-means clustering always converges to a global minimum

○ During training the number of clusters are gradually reduced until the optimal number of cluster is found.

Maximum marks: 1

16 2021-June-ML-Q16

K-fold cross validation is used for *one* of the following purposes
Select one alternative:

○ 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.

○ To avoid over fitting by reducing the dependency on one training set.

Maximum marks: 1

17 2021-ML-June-Q18

The ID3 learning algorithm for decision trees always learns an optimal decision tree
Select one alternative:

○ False

○ True

Maximum marks: 1

18 2021-ML-June-Q17

Which of the following options can be tried to get a global minima in the K-means clustering algorithm.

1. Run the algorithm for different centroid initialisations.
2. Adjust the number of iterations.
3. Adjusting the number of clusters.

Select one alternative:

○ 1

○ 1 and 3

○ 1,2,3

○ 3

Maximum marks: 1

19 2021-June-ML-Q19

What best describes the purpose of random forest algorithm for decision trees.
Select one alternative:

○ 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.

○ 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.

○ 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

20 2021 ML-Aug Q20

Suppose that you have implemented a machine learning algorithm that performs well on the training set, but make very bad predictions on a new data set. Which of the following options could you try?

Select one alternative:

○ Get more training data

○ Use a different error function.

○ Use a subset of your training data.

○ Change the learning rate parameter.

Maximum marks: 1

21 2021-ML-Naive-Bayes-a

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
1	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. Please do your best with this terrible interface to write your formulas as clearly as possible. (2 points)

Fill in your answer here

Format	B	I	U	x₂	x²	Iₓ									Ω		

Words: 0

Maximum marks: 2

22 2021-ML-Naive-Bayes-b

This is the same table as in the previous question. It is just repeated her for your convenience.

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
1	0	1	1	NO

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. (2 points)

Fill in your answer here

Words: 0

Maximum marks: 2

23 2021-ML-Decision-Trees

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 is 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

1. First calculate the entropy of the data set. You must show all your workings.
2. For each feature calculate the information gain of that feature. You must show all your workings.
3. Which node will be at the top of your decision tree and why?

Each part is worth 2 points.

Fill in your answer here

| Format ▾ | **B** | *I* | U | x_2 | x^2 | I_x | ⎗ | ⎘ | ← | → | ⟲ | ⮱ | ⋮≡ | ⋮≡ | Ω | ⊞ | ✎ |
| Σ | ✕ |

Words: 0

Maximum marks: 6

24 # 2021-ML-June-OneHot

What is one-hot encoding? Give an example and explain why it is necessary? (2 points)

Fill in your answer here

Words: 0

Maximum marks: 2