You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam 0:59:47

This section is a prerequisite. You must complete this section in order to unlock additional content.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30
31	32	33	34	35
36	37	38	39	40
41	42	43	44	45

Exam Questions

Final Exam due Aug 4, 2023 11:00 +07

Question #6fef7a

1.0 point possible (graded, results hidden)

Choose the most appropriate statement about overfitting

- ◯ A function is said to overfit relative to another one if it is less accurate in fitting known data, but more accurate in predicting unseen data
- ◯ A function is said to overfit relative to another one if it is less accurate in both fitting known data and predicting unseen data
- ◯ A function is said to overfit relative to another one if it is more accurate in fitting known data, but less accurate in predicting unseen data
- ◯ All the above mentioned statements are wrong

Submit

Question #2fbde8

1.0 point possible (graded, results hidden)

Where is the difference between supervised learning and unsupervised learning?

- ◯ From the way we train a model, supervised learning means that we have to provide detailed steps for a machine to learn
- ◯ From the training data for which supervised learning often requires labels/responses for the training phase
- ◯ From the aim of the algorithm, unsupervised learning often does not do prediction
- ◯ From the type of the output which is often a real number in supervised learning

Submit

Question #0c26ac

1.0 point possible (graded, results hidden)

Choose the most appropriate statement about underfitting

- ◯ A learning algorithm is said to underfit relative to another one if it is less accurate in

fitting known data, but less accurate in predicting unseen data

Submit

Question #949251

1.0 point possible (graded, results hidden)

Knowledge discovery is ...

- ◯ the process of creating knowledge from a dataset
- ◯ the process of grouping objects in such a way that objects in the same group are more similar than those in the other groups
- ◯ the process of converting data from one format (or structure) into a different type of format (or structure)
- ◯ the process of finding in a dataset those data points that do not have values for the input variables

Submit

Question #b1b2bc

1.0 point possible (graded, results hidden)

Which of the following is a solution to cleaning noise data?

- ☐ Binning
- ☐ Outlier detection
- ☐ Normalization
- ☐ Aggregation

Submit

Question #c96b71

1.0 point possible (graded, results hidden)

Data discretization is

- ◯ to convert discrete attributes to continuous ones
- ◯ to find outliers
- ◯ to convert continuous attributes to discrete ones
- ◯ to scale up data

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

1.0 point possible (graded, results hidden)

Data preprocessing may refer to

- ◯ deciding the data format and collecting data then
- ◯ process of transforming a raw dataset into a suitable form for analysis
- ◯ analyzing some main properties of the dataset

Submit

Question #5255e8

1.0 point possible (graded, results hidden)

What does "mode" mean?

- ◯ Variation
- ◯ Repeats the most
- ◯ Middle
- ◯ Average

Submit

Question #e22c51

1.0 point possible (graded, results hidden)

Normalization is the process of

- ◯ replacing missing values
- ◯ removing outliers
- ◯ transforming the data from one vector space to another one
- ◯ transforming the data to fall within a common range

Submit

Question #532c0d

1.0 point possible (graded, results hidden)

What is the role of an loss/error function in machine learning?

- ◯ No role in the machine learning process
- ◯ To measure the error in some senses and to often play as the objective function for training a model
- ◯ To measure the loss/error when making future prediction

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

Question #e6aa98

1.0 point possible (graded, results hidden)

What is the wrong statement in the followings?

- ◯ Model assessment and model selection in machine learning are independent
- ◯ Model selection is a must when comparing different machine learning models/methods
- ◯ Model assessment often requires model selection as an internal step

Submit

Question #b2a230

1.0 point possible (graded, results hidden)

What is the drawback of Hold-out, but can be overcome by stratified sampling for evaluation?

- ◯ The bad effect of randomness on evaluation results, due to small data size
- ◯ The bad effect of the imbalance between classes
- ◯ The bad effect of evaluation time

Submit

Question #89a2ec

1.0 point possible (graded, results hidden)

Is Hold-out a method for data pre-processing and understanding?

- ◯ No, it is a strategy for model assessment and selection
- ◯ Yes, of course
- ◯ No, it is a method for training a model from a given dataset

Submit

Question #4f3e62

1.0 point possible (graded, results hidden)

Your work needs to build a classifier that can classify a given email to be spam or normal. However you can only collect a dataset with severe imbalance, that is, 99.9% of the emails are spam. What measure should you use to evaluate the performance of your classifier?

- ☐ Accuracy
- ☐ Precision for individual class

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

Submit

Question #9f0a3e

1.0 point possible (graded, results hidden)

The ordinary least squares (OLS) method learns a function $f(x) = w_0 + w_1 x_1 + \cdots + w_n x_n$ from a dataset of size M by minimizing the loss $L = \sum_{i=1}^{M} (y_i - w_0 - w_1 x_{i1} - \cdots - w_n x_{in})^2$. Consider the regularized loss $L_\lambda = \sum_{i=1}^{M} (y_i - w_0 - w_1 x_{i1} - \cdots - w_n x_{in})^2 + \lambda \|\mathbf{w}\|_2^2$, where λ is a non-negative constant. Which of the following statement is WRONG?

- λ can preserve the goodness of the solution of OLS
- λ can help to reducing overfitting
- λ plays as the central role on the generalization of $f(x)$

Submit

Question #a9c1ae

1.0 point possible (graded, results hidden)

What kind of models does the following regression model belong to?
$f(x) = w_0 + w_1 x_1 + \cdots + w_n x_n$ where $w_0, w_1, \ldots, w_n$ are the regression coefficients

- A non-linear model
- A non-parametric model
- A linear model

Submit

Question #f4bde8

1.0 point possible (graded, results hidden)

The least-square linear regression method is applicable only in situations where the estimated regression line has a positive slope/bias.

- No
- Yes

Submit

Question #4193c1

1.0 point possible (graded, results hidden)

They are the same

They are overlapping

It is inappropriate to say that they are overlapping or disjoint

They are disjoint

Submit

Question #c6c360

1.0 point possible (graded, results hidden)

What can happen when we use an extremely large number of nearest neighbors for prediction in k-NN?

The prediction will be prone to trivial

The prediction will tend to be more accurate

k-NN will tend to get under-fitting

k-NN will tend to get over-fitting

Submit

Question #96339b

1.0 point possible (graded, results hidden)

Which of the following options is true about Nearest Neighbor-based Learning (k-NN) algorithm?

It can be used for both classification and regression

It can be used for regression only

It can be used for classification only

Submit

Question #0a4211

1.0 point possible (graded, results hidden)

Which of the following statements is true for Manhattan distance function?

It can be used for continuous variables

It can be used for nominal variables

It can be used for neither continuous nor nominal variables

It can be used for both continuous and nominal variables

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

1.0 point possible (graded, results hidden)

Which of the following values is the Euclidean distance between the two data points X(1,7) and Y(4,3)?

- ◯ 3
- ◯ 1
- ◯ 9
- ◯ 7
- ◯ 5

Submit

Question #bc8473

1.0 point possible (graded, results hidden)

Which of the following statements are true about the Nearest Neighbor-based Learning (k-NN) algorithm?
A. In case of very large value of k, it may include data points from other classes into the neighborhood.
B. In case of too small value of k, it is very sensitive to noise.

- ◯ A and B
- ◯ A
- ◯ Neither A nor B
- ◯ B

Submit

Question #b9aede

1.0 point possible (graded, results hidden)

x_1	x_2	Class
-1	1	c_2
0	1	c_1
0	2	c_2
1	-1	c_2
1	0	c_1
1	2	c_1
2	2	c_2
2	3	c_1

Using the Nearest Neighbor-based Learning (k-NN) algorithm and Euclidean distance and 7 nearest neighbors, which class does the data point ($x_1 = 1, x_2 = 1$) belong to?

- ◯ c_2
- ◯ Undefined
- ◯ c_1

Submit

Question #fc7a08

1.0 point possible (graded, results hidden)

Maximum likelihood estimation can be used to

- ☐ make inference/prediction for a new example/observation
- ☐ explore the knowledge in a learned model
- ☐ estimate the maximum likelihood of a model
- ☐ learn a model from a given training dataset
- ☐ infer the correctness of a given model

Submit

Question #265f13

1.0 point possible (graded, results hidden)

Posterior probability may refer to

- ◯ the knowledge of our model
- ◯ probability of an observation given a model/hypothesis
- ◯ probability of a model/hypothesis given observed data

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

1.0 point possible (graded, results hidden)

What are wrong about Maximum A Posteriori (MAP) estimation?

- [ ] MAP needs not to know any knowledge about the parameters of a model
- [ ] MAP needs to know some knowledge about the parameters of a model
- [ ] MAP can estimate the full posterior distribution from a training dataset
- [ ] MAP can be used to learn a model from a training dataset
- [ ] MAP can make inference or prediction for new examples

Submit

Question #04be97

1.0 point possible (graded, results hidden)

The ID3 decision tree learning algorithm can be used for ...

- ◯ regression
- ◯ both classification and regression
- ◯ neither classification nor regression
- ◯ classification

Submit

Question #a23c56

1.0 point possible (graded, results hidden)

Decision tree is ... algorithm.

- ◯ a supervised learning
- ◯ both a supervised and an unsupervised learning
- ◯ neither a supervised nor an unsupervised learning
- ◯ an unsupervised learning

Submit

Question #27df10

1.0 point possible (graded, results hidden)

Given a learned decision tree, a test example's class label is determined by ... in the tree.

- ◯ following a single path
- ◯ following a path with minimum depth

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

Submit

Question #dc8ba6

1.0 point possible (graded, results hidden)

Over-fitting in Decision Tree can be handled by ...

- [ ] Reducing the number of input attributes
- [ ] Reducing the number of training examples
- [ ] Stopping the growing of the decision tree earlier
- [ ] Translating to a set of rules and pruning the rules
- [ ] Growing a complete tree

Submit

Question #16ad1d

1.0 point possible (graded, results hidden)

Which of the following are true about decision trees?

- ◯ They can be used only for regression.
- ◯ Pruning usually achieves better test accuracy than stopping early.
- ◯ All data instances in each leaf must have the same class.

Submit

Question #3a5111

1.0 point possible (graded, results hidden)

Which of the following are true of neural networks?

- [ ] Optimize a linear objective function
- [ ] Can only be trained with stochastic gradient descent
- [ ] Can use a mix of different activation functions
- [ ] Can be made to perform well even when the number of parameters/weights is much greater than the number of data points

Submit

Question #c42a8e

1.0 point possible (graded, results hidden)

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

Submit

Question #54d929

1.0 point possible (graded, results hidden)

What is activation function?

- ◯ It computes the output of the neural network
- ◯ It computes weighted sum of a neuron's input signals
- ◯ It computes the output of a neuron given its net input
- ◯ It computes the output of a neuron given its input signals

Submit

Question #59f1a9

1.0 point possible (graded, results hidden)

A neural network can be seen as a non-linear function. Which is/are that function non-linear in?

- ◯ Both input signal x and weights
- ◯ The weights
- ◯ Input signal x

Submit

Question #3aa9b4

1.0 point possible (graded, results hidden)

When training a neural network, we are trying to search for ...

- ☐ A function in the function family, which maps from an example to an output
- ☐ The best weights from the weight space
- ☐ The hyper-parameter(s) that give best performance
- ☐ An estimate of the testing error
- ☐ A configuration of the network architecture

Submit

- [ ] always converges to a clustering that minimizes the mean-square vector-representative distance
- [ ] can converge to different final clustering, depending on initial choice of clusters' centroids
- [ ] is sensitive to outliers

Submit

Question #ab0712

1.0 point possible (graded, results hidden)

A cluster is

- ◯ A group of similar objects that differ significantly from other objects in a different group
- ◯ Symbolic representation of facts or ideas from which information can potentially be extracted
- ◯ A group of similar or different objects
- ◯ Operation on a database to transform or simplify data in order to prepare it for a machine learning or data mining algorithm

Submit

Question #8d15ad

1.0 point possible (graded, results hidden)

Consider the following dataset: $A = (0; 2)$, $B = (0; 1)$ and $C = (1; 0)$. The K-means algorithm (with the Euclidean distance) is initialized with centers at A and B. Upon convergence, the two centers will be at

- ◯ C and the midpoint of AB
- ◯ A and B
- ◯ A and C
- ◯ A and the midpoint of BC

Submit

Question #41dbd9

1.0 point possible (graded, results hidden)

The K-means algorithm ...

- ◯ requires the dimension of the feature space to be no bigger than the number of samples

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

- ◯ minimizes the within cluster variance for a given number of clusters

Submit

Question #31c916

1.0 point possible (graded, results hidden)

You want to cluster 7 points into 3 clusters using the k-Means clustering algorithm (with the Euclidean distance). Suppose after the first iteration, clusters C1, C2 and C3 contain the following two-dimensional points: C1 contains the 2 points {(0,6), (6,0)}, C2 contains the 3 points {(2,2), (4,4), (6,6)}, and C3 contains the 2 points {(5,5), (7,7)} What are the cluster centers computed for these 3 clusters?

- ◯ C1: (3,3), C2: (4,4), C3: (6,6)
- ◯ C1: (6,6), C2: (12,12), C3: (12,12)
- ◯ C1: (0,0), C2: (48,48), C3: (35,35)
- ◯ C1: (3,3), C2: (6,6), C3: (12,12)

Submit

Question #9bc723

1.0 point possible (graded, results hidden)

What does the Apriori algorithm do?

- ☐ prune rules whose support are lower than minimum support (minsup)
- ☐ prune rules whose sup are higher than minimum support (minsup)
- ☐ generate rules whose confidence are lower than minimum confidence (minconf)
- ☐ generate rules whose confidence are higher than minimum confidence (minconf)

Submit

Question #10315e

1.0 point possible (graded, results hidden)

What does it mean by support count of itemset A?

- ◯ Total number of transactions not containing A
- ◯ Number of transactions not containing A / Total number of transactions
- ◯ Number of transactions containing A / Total number of transactions
- ◯ Total number of transactions containing A

You are taking "Exam on August 04, 2023" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam".

0:59:47

1.0 point possible (graded, results hidden)

Consider a set of 6 transactions as shown in the following table.

Transaction ID	*Bought items*
T1	{A, B, E}