In a linear model, W, b, X, y are known as:

- Bias, Weight, Input, Output
- Bias, Weight, Output, Input
- Weight, Bias, Input, Output
- Bias, Output, Weight, Input

As the level of recall rises, the level of precision generally:

- Declines
- Rises too
- Stays the same

Linear regression, mathematically, is given by:

- $\sum(b + X)$
- $\sum(X + W)$
- $\sum(W*X + b)$
- $\sum(W*b + X)$

Which of the following metrics used to evaluate the performance of linear regression

- F1 score
- Accuracy
- Mean Square Error

Overfitting is when the model is too complex and:

- Training error is large while test error is small
- Both training and test errors are small
- Both training and test errors are large
- Training error is small while test error is large

Which is the correct way to represent Multi-label Classification:

- Each label belongs to one or more {1,2,...,M}
- Each label belongs to one or more {-1,1}
- Each label belongs to either {1,2,...,M}
- Each Label belongs to exactly either {-1,1}

Using L0 norm encourages very few:

- Zeroes in W
- Non-zero entries in W
- Error in W

Supervised learning requires the following:

- [x] Input-output pairs
- [ ] Only outputs
- [ ] Only inputs

In cross-validation, training data is divided into:

- Training and test sets
- Training and validation sets
- Training, validation, and optimization sets
- Training, validation, and test sets

Choose the correct pair, Very small and very large regularization can cause:

- Underfitting and fitting
- fitting and overfitting
- Underfitting and Overfitting
- Overfitting and Underfitting

Performance on test data is a good indicator of generalization performance

- ○ Incorrect
- ✓ Correct

L1 regularization is also useful for doing:

- [x] Feature selection
- [ ] Regression
- [ ] Feature extraction

Which of these is a categorical feature:

- male/female
- Pixel intensity
- Blood-group
- Grade in a course

Which is the correct way to represent Binary classification:

- y (belongs to){-1,1}
- y (belongs to){1,2,...,M}
- y (belongs to) R (to the power of M)

Underfitting is when the model is too simple and:

- Training error is large while test error is small
- ✓ Both training and test errors are large
- Both training and test errors are small
- Training error is small while test error is large

The difference between Regression and Multiple-Output Regression is:

- In regression output is a label, whereas Multiple-Output Regression output is a real value scalar
- In regression output is a vector containing M output, whereas Multiple-Output Regression output is real value scalar
- In regression output is a real value scalar, whereas Multiple-Output Regression output is a vector containing M output
- In regression output is a real value scalar, whereas Multiple-Output Regression output is a label

Question 1 10 out of 10 points

Using L2 norm in the loss function promotes the individual entries of Weights (W) to be:

Question 2 10 out of 10 points

In a linear model, W, b, X, y are known as:

Question 3 10 out of 10 points

As the level of recall rises, the level of precision generally:

Question 4 10 out of 10 points

Linear regression, mathematically, is given by:

Question 5 10 out of 10 points

Which of these is not a binary feature:

Question 6 10 out of 10 points

Which of the following metrics used to evaluate the performance of linear regression

Question 7 10 out of 10 points

Overfitting is when the model is too complex and:

Question 8 0 out of 10 points

Which is the correct way to represent Multi-label Classification:

Question 9 10 out of 10 points

Using L0 norm encourages very few:

Question 10 10 out of 10 points

Supervised learning requires the following:

Question 11 10 out of 10 points

In cross-validation, training data is divided into:

Question 12 10 out of 10 points

Choose the correct pair, Very small and very large regularization can cause:

Question 13 10 out of 10 points

Performance on test data is a good indicator of generalization performance

Question 14 10 out of 10 points

L1 regularization is also useful for doing:

Question 15 10 out of 10 points

Which of these is a categorical feature:

Question 16 10 out of 10 points

Which is the correct way to represent Binary classification:

Question 17 10 out of 10 points

Underfitting is when the model is too simple and:

Question 18 10 out of 10 points

The difference between Regression and Multiple-Output Regression is:

In cross-validation, training data is divided into:

- Training and test sets
- Training and validation sets
- Training, validation, and optimization sets
- Training, validation, and test sets

Remaining Time: 1 hour, 26 minutes, 46 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 6

The difference between Regression and Multiple-Output Regression is:

- In regression output is a label, whereas Multiple-Output Regression output is a real value scalar
- In regression output is a real value scalar, whereas Multiple-Output Regression output is a vector containing M output
- In regression output is a real value scalar, whereas Multiple-Output Regression output is a label
- In regression output is a vector containing M output, whereas Multiple-Output Regression output is real value scalar

Moving to the next question prevents changes to this answer.

n 1

Midterm 1 (Part 2)

Part 2 is available. Please submit it before the deadline! (Sat 11:59 PM)

Multiple Attempts	Not allowed. This test can only be taken once.
Force Completion	Once started, this test must be completed in one sitting. Do not leave the test before clicking **Save and Submit**.
	This test does not allow backtracking. Changes to the answer after submission are prohibited.

Remaining Time: 1 hour, 29 minutes, 36 seconds.

Question Completion Status:

⚠ Moving to the next question prevents changes to this answer.

Question **1** of **12**

Question 1

10 points Save Answer

Which of these is false about Decision Tree (DT):

- It can easily handle different types of features
- It is simple and easy to interpret
- Very fast at test time
- ✔ Multiple DTs cannot be combined via ensemble methods

⚠ Moving to the next question prevents changes to this answer.

Question **1** of **12**

This test does not allow backtracking. Changes to the answer after submission are prohibited.

Remaining Time: **1 hour, 24 minutes, 52 seconds.**

Question Completion Status:

⚠ Moving to the next question prevents changes to this answer.

Question **2** of **12**

Question 2

10 points Save Answer

Which of the following is true about the nearest neighbors algorithm:

- ✔ Prediction can be slow at test time
- ○ 1-NN doesn't suffer from outliers
- ○ Requires a small amount of storage

⚠ Moving to the next question prevents changes to this answer.

Question **2** of **12**

Remaining Time: **1 hour, 20 minutes, 52 seconds.**

Question Completion Status:

⚠ Moving to the next question prevents changes to this answer.

Question **3** of **12**

Question 3

10 points Saved

Which of the following is false about k-NN:

- Large K leads to non-smooth boundaries
- The decision boundary is composed of hyperplanes
- k-NN is geometry-based
- k-NN is a distance-based algorithm

⚠ Moving to the next question prevents changes to this answer.

Question **3** of **12**

Choose the correct pair, one of the limitations of the most classification algorithm is that:

- Assumes classes are equally sized and requires small training dataset
- ✔ Assumes classes are equally sized and requires large training dataset
- Assumes classes are not equally sized and requires small training dataset
- Assumes classes are not equally sized and requires large training dataset

Choose the correct pair, one of the limitations of the most classification algorithm is that:

- Assumes classes are equally sized and requires small training dataset
- Assumes classes are equally sized and requires large training dataset
- Assumes classes are not equally sized and requires small training dataset
- Assumes classes are not equally sized and requires large training dataset

Decision Trees can also be used for regression problems

- [ ] No
- [x] Yes

In k-NN, very small and very large values of K can lead to:

- Overfitting
- Underfitting and overfitting
- Overfitting and underfitting
- Underfitting

Question 2

10 points | Save Answer

In the nearest neighbor algorithm, the averaging version of prediction rule is given by

- y = 1/k ∑Yn
- y= k ∑Xn
- y= 1/k ∑Xn
- y= k ∑Yn

⚠ Moving to the next question prevents changes to this answer.

Question **2** of **12**

Question 2

10 points Save Answer

In the nearest neighbor algorithm, the averaging version of prediction rule is given by

- $y = 1/k \sum Yn$
- $y = k \sum Xn$
- $y = 1/k \sum Xn$
- $y = k \sum Yn$

⚠ Moving to the next question prevents changes to this answer.

Question **2** of **12**

One of the key properties of Decision-Trees is:

- Prediction at train time is very fast
- Prediction at test time is very fast
- Prediction at test time is slow

Question 3

10 points Save Answer

Which of the following is false about k-NN:

- ○ The decision boundary is composed of hyperplanes
- ○ k-NN is a distance-based algorithm
- ○ k-NN is geometry-based
- ✓ Large K leads to non-smooth boundaries

⚠ Moving to the next question prevents changes to this answer.

Question **3** of **12**

Which of these is false about Decision Tree (DT):

- Multiple DTs cannot be combined via ensemble methods
- It is simple and easy to interpret
- It can easily handle different types of features
- Very fast at test time

Which of the following is **false**, stop expanding Decision Trees when:

- ○ Run out of features to test along the path to that node
- ✓ Run out of training examples
- ○ It consists of examples all having the same label

⚠ Moving to the next question prevents changes to this answer.

Which of the following is true about the nearest neighbors algorithm:

- Requires a small amount of storage
- Prediction can be slow at test time
- 1-NN doesn't suffer from outliers

Which of the following is false about k-NN:

- The decision boundary is composed of hyperplanes
- k-NN is a distance-based algorithm
- k-NN is geometry-based
- Large K leads to non-smooth boundaries

Select the correct pair(for decision tree):

- ◯ Root, internal and leaf node contain rules
- ◉ Root and internal node contain rule, leaf node contain a prediction
- ◯ Root, internal and leaf node contain predictions
- ◯ Root and internal node contain prediction, leaf node contain rule

At the internal nodes in Decision Tress, splits should be such that:

- [x] Low entropy, Non-uniform label
- [ ] High entropy, uniform label
- [ ] High entropy, non-uniform label
- [ ] Low entropy, uniform label

Question 2

Which of these is false about Decision Tree (DT):

- It can easily handle different types of features
- It is simple and easy to interpret
- Multiple DTs cannot be combined via ensemble methods
- Very fast at test time

⚠ Moving to the next question prevents changes to this answer.

Which of the following is **false**, stop expanding Decision Trees when:

- [ ] Run out of training examples
- [ ] It consists of examples all having the same label
- [ ] Run out of features to test along the path to that node

Question Completion Status:

Moving to another question will save this response.

Question 1

Bagging ensemble is useful for models with

- [x] High variance and noisy data
- [ ] High variance and clean data
- [ ] Low variance and clean data

Moving to another question will save this response.

⚠ Moving to another question will save this response.

Question 1

Let's assume that we are building a Decision Tree (DT), At the root node (S) we have a feature with: (Rain = 15, no rain= 10)

Entropy: H(S) is equal:

- [ ] 0.92
- [ ] 0.95
- [ ] 0.9
- [ ] 0.88

⚠ Moving to another question will save this response.

Select true statement:

- [ ] Complex models have a higher bias. Boosting can reduce it
- [ ] Complex models have a higher bias. Bagging can reduce it
- [x] Complex models have higher variance. Bagging can reduce it
- [ ] Simple models have a higher bias. Bagging can reduce it
- [x] Simple models have a higher bias. Boosting can reduce it
- [ ] Complex models have higher variance. Boosting can reduce it
- [ ] Simple models have higher variance. Bagging can reduce it
- [ ] Simple models have higher variance. Boosting can reduce it

Question 1

Which of the following statements are true for Bagging and Boost training:

- [ ] Bagging trains independent models sequentially
- [ ] Boosting trains independent models
- [x] Boosting trains models sequentially
- [ ] Bagging trains models sequentially
- [ ] Boosting trains sequential model independently
- [x] Bagging trains independent models

⚠ Moving to another question will save this response.

Question 1

For classification problems, the entropy of the label distribution is a measure of:

- [ ] Randomness
- [ ] Noise
- [x] Purity
- [ ] Impurity

⚠ Moving to another question will save this response.

Question 2

In kNN, the decision boundary is:

- [ ] Composed of non-linear hyperplanes
- [x] Composed of hyperplanes
- [ ] Composed of linear boundaries
- [ ] Composed of linear lines

ning

⚠ Moving to another question will save this response.

Question 2

Select true statement:

- [ ] Simple models have higher variance. Boosting can reduce it
- [ ] Complex models have a higher bias. Boosting can reduce it
- [ ] Simple models have a higher bias. Bagging can reduce it
- [x] Simple models have a higher bias. Boosting can reduce it
- [ ] Complex models have a higher bias. Bagging can reduce it
- [x] Complex models have higher variance. Bagging can reduce it
- [ ] Complex models have higher variance. Boosting can reduce it
- [ ] Simple models have higher variance. Bagging can reduce it

Question 3

Bagging creates training data by

- [ ] Create from the original dataset with sampling and replacement over-weighted data
- [x] Create a copy from the original dataset with random sampling and replacement
- [ ] Extract the same data from the original dataset

Question Completion Status:

Moving to another question will save this response.

Question 2

2 p

Which of these statements are true about stacking:

- [ ] Average the predictions of multiple models on the test data
- [ ] Use predictions of multiple models as features to test a new model and use the new model to train data
- [ ] Use features of multiple models to train a new model and use the new model to make predictions on test data
- [x] Use predictions of multiple models as features to train a new model and use the new model to make predictions on test data

Moving to another question will save this response.

Question 1

Random Forest is a special case of:

- [ ] Classification
- [ ] Linear Models
- [x] Bagging
- [ ] Boosting

Projects

Midterm 1

Final

Question 3

2 points Saved

The general steps for training algorithms in Boosting are:

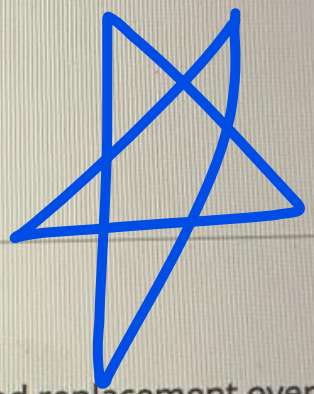
- [x] 1) Train a weak model
  2) Compute the error of the model
  3) Give higher importance to examples on which the model made mistakes
  4) Re-train the model
  5) Go back to step 2

- [ ] 1) Train a weak model
  2) Compute the error of the model
  3) Give lower importance to examples on which the model made mistakes
  4) Re-train the model
  5) Go back to step 2

- [ ] 1) Train a strong model
  2) Compute the error of the model
  3) Give higher importance to examples on which the model made mistakes
  4) Re-train the model
  5) Go back to step 2

- [ ] 1) Train a strong model
  2) Compute the error of the model
  3) Give a lower importance to examples on which the model made mistakes
  4) Re-train the model
  5) Go back to step 2

⚠ Moving to another question will save this response.



Question 2

Boosting creates training data by:

- [ ] Creates copy from original data set with random sampling and replacement over-weighted data
- [ ] Creates copy from original data set with random sampling and replacement
- [ ] Extract the same data from the original dataset

⚠ Moving to another question will save this response.

Question Completion Status:

Moving to another question will save this response.

Question 3

K-means loss function assumes:

- [ ] Clusters are not equal
- [x] Clusters are equal-sized
- [ ] CLusters are random shaped

Moving to another question will save this response.

Question completion status:

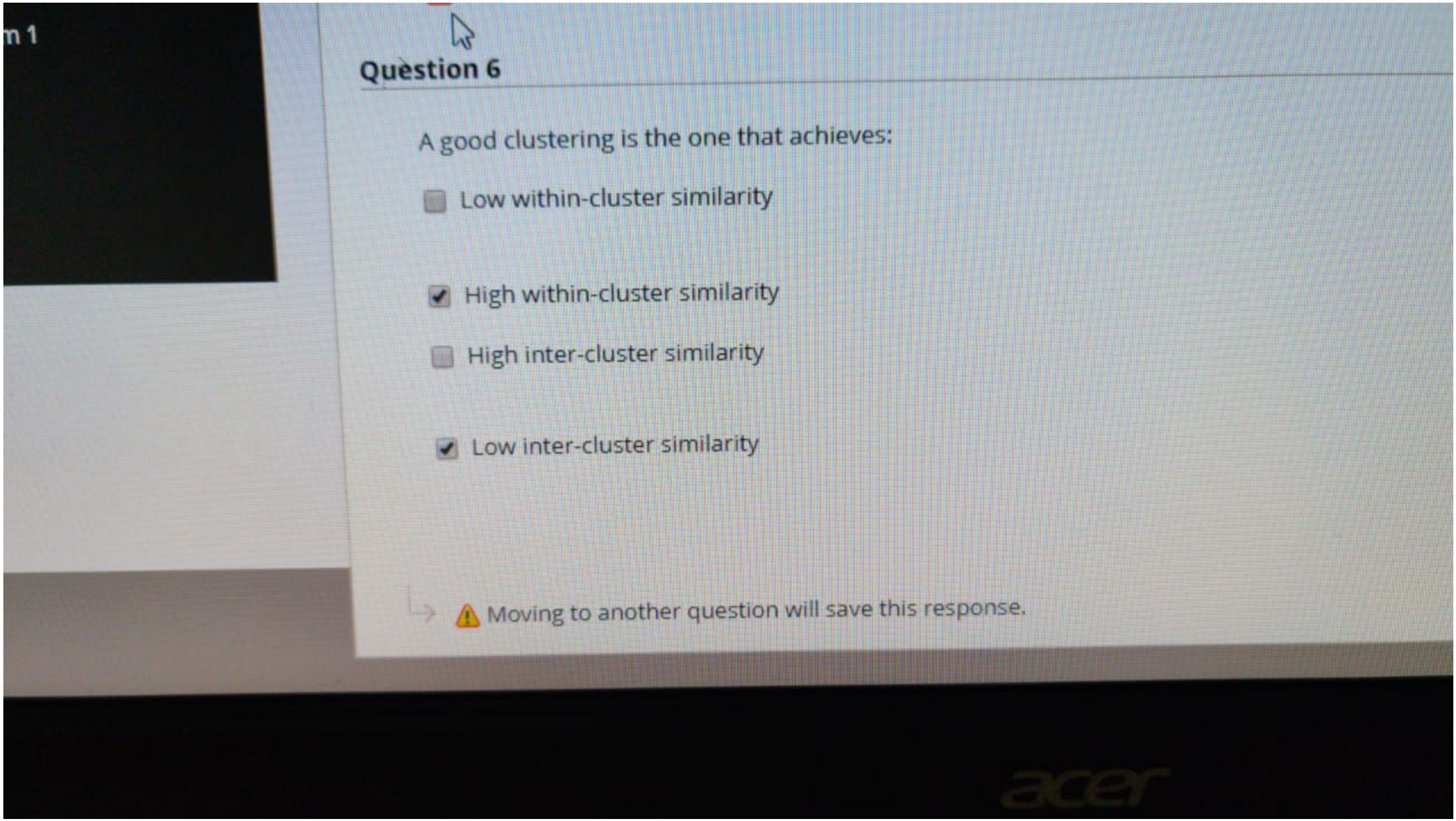
⚠ Moving to another question will save this response.

Question 5

A good clustering is the one that achieves:

- [ ] High inter-cluster similarity
- [x] High within-cluster similarity
- [x] Low inter-cluster similarity
- [ ] Low within-cluster similarity

⚠ Moving to another question will save this response.


m 1
Question 6
A good clustering is the one that achieves:
Low within-cluster similarity
High within-cluster similarity
High inter-cluster similarity
Low inter-cluster similarity
Moving to another question will save this response.
acer

⚠ Moving to another question will save this response.

Question 7

Which of the following statements are true for Bagging and Boost training:

- [x] Boosting trains models sequentially
- [ ] Bagging trains models sequentially
- [x] Bagging trains independent models
- [ ] Boosting trains independent models
- [ ] Bagging trains independent models sequentially
- [ ] Boosting trains sequential model independently

⚠ Moving to another question will save this response.

Boosting creates training data by:

- [ ] Creates copy from original data set with random sampling and replacement over-weighted data
- [ ] Extract the same data from the original dataset
- [ ] Creates copy from original data set with random sampling and replacement

⚠ Moving to another question will save this response.

Question 9

Random Forest is a special case of:

- [x] Bagging
- [ ] Linear Models
- [ ] Classification
- [ ] Boosting

⚠ Moving to another question will save this response.

⚠ Moving to another question will save this response.

Question 11

In Bagging, the final model is:

- [ ] Minimum model
- [ ] Maximum model
- [x] Averaged model

⚠ Moving to another question will save this response.

Which of these statements are true about stacking:

- [x] Use predictions of multiple models as features to train a new model and use the new model to make predictions on test data
- [ ] Average the predictions of multiple models on the test data
- [ ] Use features of multiple models to train a new model and use the new model to make predictions on test data
- [ ] Use predictions of multiple models as features to test a new model and use the new model to train data

Question 7

Which of the following statements are true for Bagging and Boost training:

- [ ] Bagging trains independent models sequentially
- [ ] Boosting trains independent models
- [ ] Bagging trains independent models
- [ ] Boosting trains sequential model independently
- [ ] Bagging trains models sequentially
- [ ] Boosting trains models sequentially

⚠ Moving to another question will save this response.

Question 13

Bagging creates training data by

- [ ] Extract the same data from the original dataset
- [ ] Create from the original dataset with sampling and replacement over-weighted data
- [ ] Create a copy from the original dataset with random sampling and replacement

⚠ Moving to another question will save this response.

⚠ Moving to another question will save this response.

Question 2

Boosting creates training data by:

- [ ] Creates copy from original data set with random sampling and replacement over-weighted data
- [ ] Creates copy from original data set with random sampling and replacement
- [ ] Extract the same data from the original dataset

⚠ Moving to another question will save this response.

earning

⚠ Moving to another question will save this response.

Question 13

Bagging ensemble is useful for models with

- [ ] Low variance and clean data
- [ ] High variance and clean data
- [x] High variance and noisy data

⚠ Moving to another question will save this response.

Question 8

Select true statement:

- [ ] Simple models have a higher bias. Boosting can reduce it
- [ ] Simple models have higher variance. Boosting can reduce it
- [ ] Simple models have a higher bias. Bagging can reduce it
- [ ] Complex models have higher variance. Boosting can reduce it
- [ ] Complex models have a higher bias. Bagging can reduce it
- [ ] Complex models have higher variance. Bagging can reduce it
- [ ] Complex models have a higher bias. Boosting can reduce it
- [ ] Simple models have higher variance. Bagging can reduce it

K-means loss function assumes:

- [ ] Clusters are not equal
- [x] Clusters are equal-sized
- [ ] CLusters are random shaped

⚠ Moving to another question will save this response.

Boosting creates training data by:

- [ ] Creates copy from original data set with random sampling and replacement
- [ ] Creates copy from original data set with random sampling and replacement over-weighted data
- [ ] Extract the same data from the original dataset

Question 7

Which of the following statements are true for Bagging and Boost training:

- [x] Boosting trains models sequentially
- [ ] Bagging trains models sequentially
- [x] Bagging trains independent models
- [ ] Boosting trains independent models
- [ ] Bagging trains independent models sequentially
- [ ] Boosting trains sequential model independently

⚠ Moving to another question will save this response.

Question 10

For classification problems, the entropy of the label distribution is a measure of:

- [ ] Randomness
- [ ] Noise
- [ ] Purity
- [ ] Impurity

Question 12

In Bagging, the final model is:

- Averaged model
- Maximum model
- Minimum model

Question 1

Select true statement:

- [ ] Complex models have a higher bias. Boosting can reduce it
- [ ] Complex models have higher variance. Bagging can reduce it
- [ ] Simple models have a higher bias. Bagging can reduce it
- [ ] Simple models have higher variance. Boosting can reduce it
- [x] Simple models have a higher bias. Boosting can reduce it
- [ ] Complex models have higher variance. Boosting can reduce it
- [ ] Complex models have a higher bias. Bagging can reduce it
- [ ] Simple models have higher variance. Bagging can reduce it

Select true statement:

- [ ] Complex models have a higher bias. Boosting can reduce it
- [ ] Complex models have a higher bias. Bagging can reduce it
- [x] Complex models have higher variance. Bagging can reduce it
- [ ] Simple models have a higher bias. Bagging can reduce it
- [x] Simple models have a higher bias. Boosting can reduce it
- [ ] Complex models have higher variance. Boosting can reduce it
- [ ] Simple models have higher variance. Bagging can reduce it
- [ ] Simple models have higher variance. Boosting can reduce it

Hi, a couple of students asked about the content of the final exam. I thought to provide a full answer for all of you:

- The final exam will cover everything we covered during the lectures and tutorials.

- The main goal of this course is to learn and understand the core concept of machine learning algorithms and their practical applications. You will not be tested on any mathematical concept, especially optimization.

- The exam questions will be only in the form of **multiple choice**. There will be no code writing. However, assignments and their solutions might be included in the final exams as multiple answer questions.