# Q1 Exam regulations

0 Points

**BBM406 Honor Code**

I promise that, for the BBM406 Final Exam

- All my exam work will be done entirely by myself, with no help from others;
- I will not communicate with anybody except the proctors during the exam;
- I will not consult any people or sources other than my printed/handwritten course notes, slides and the reference books listed in the course webpage;
- I will not provide any information about the exam's contents to other students until the exam deadline; and
- I will turn on my camera on Zoom session during the whole exam period.

*Understanding this, I pledge my honor that I will not violate this Honor Code during the exam. I certify that all solutions will be entirely my own, that I will not consult people or sources other than those permitted, and that I will not share information with others during the exam.*

*Do NOT sign nor take this exam if you do not agree with this.*

Signature (Specify your name and surname as your signature)

MTUSTA Mehmet Taha USTA

# Q2

3 Points

Assume that your training set consists of the following training samples (2D feature vectors) and the associated class labels.

| $x$ | $y$ | Class |
|---|---|---|
| 0 | 2 | Negative |

| $x$ | $y$ | Class |
|---|---|---|
| 1 | 2 | Positive |
| -1 | 1 | Negative |
| 0 | 1 | Positive |
| 1 | -1 | Negative |
| 1 | 0 | Positive |
| 2 | 3 | Positive |
| 2 | 2 | Negative |

## Q2.1
1 Point

How does a 3-NN classifier classify the test point (1,1)?

- (●) Positive
- ( ) Negative

## Q2.2
1 Point

How does a 5-NN classifier classify the test point (1,1)?

- (●) Positive
- ( ) Negative

## Q2.3
1 Point

How does a 7-NN classifier classify the test point (1,1)?

○ Positive

◉ Negative

## Q3
18 Points

For each statement below, state whether it is True or False.

### Q3.1
1 Point

Mapping the original features into another feature space via a radial basis kernel function might make a perceptron achieve better classification performance as compared to using the original features.

○ True

◉ False

### Q3.2
1 Point

Mapping the original features into another feature space via a radial basis kernel function might make a 1-NN classifier achieve better classification performance as compared to using the original features (assuming unweighted Euclidean distance is used).

◉ True

○ False

### Q3.3
1 Point

Increasing training data always decreases the risk of overfitting.

- [x] True
- [ ] False

## Q3.4
1 Point

For small values of k, k-means algorithm and k-NN algorithm lead to the same results.

- [ ] True
- [x] False

## Q3.5
1 Point

Selecting the hyperparameters of a learning model through cross validation always reduces overfitting.

- [x] True
- [ ] False

## Q3.6
1 Point

An ensemble of learning models always has more capacity as compared to a single learning model.

- [ ] True
- [x] False

## Q3.7
1 Point

Logistic regression will always find the same decision boundary as a linear SVM.

○ True

◉ False

## Q3.8
1 Point

For noisy training data, random forests usually achieve better classification performance as compared to AdaBoost.

◉ True

○ False

## Q3.9
1 Point

In Adaboost algorithm, the ratio of misclassified examples to the total number of examples gives the error of each hypothesis.

○ True

◉ False

## Q3.10
1 Point

A random forest is an ensemble machine learning model which aims at decreasing the bias error of the decision trees.

◉ True

○ False

## Q3.11
1 Point

As model complexity increases, variance will increase while bias will decrease.

- [x] True
- [ ] False

## Q3.12
1 Point

The classifiers computed by kernel methods can be regarded as a linear function of its parameters but not necessarily a linear function of the input features.

- [ ] True
- [x] False

## Q3.13
1 Point

Convolutional neural networks are translation invariant.

- [ ] True
- [x] False

## Q3.14
1 Point

Increasing the depth of a decision tree will always guarantee a better training accuracy but it might reduce its robustness.

- [x] True
- [ ] False

## Q3.15
1 Point

An artificial neural network without any hidden units and trained via cross-entropy loss is equivalent to logistic regression.

○ True

◉ False

## Q3.16

1 Point

An artificial neural network having multiple hidden layers with sigmoid activation functions can lead to non-linear decision boundaries.

◉ True

○ False

## Q3.17

1 Point

Hierarchical clustering models require to set the number of clusters predefined.

○ True

◉ False

## Q3.18

1 Point

Finding the global optimum solution through k-means method is NP-hard, given a predefined number of clusters k.

◉ True

○ False

## Q4

1 Point

For an artificial neural network, which one of the following factors affects the trade-off between overfitting and underfitting the most?

- ○ Learning rate
- ◉ Initial parameter weights
- ○ Number of hidden nodes
- ○ None of the above

## Q5
1 Point

For polynomial regression, which one of the following factors affects the trade-off between overfitting and underfitting the most?

- ○ Degree of the polynomial
- ○ Obtaining model parameters via the closed-form solution or Gradient Descent
- ◉ Amount of noise in the training data
- ○ Including the bias to the input

## Q6
1 Point

For a Gaussian Bayes classifier, which one of the following factors affects the trade-off between overfitting and underfitting the most?

- ○ Learning the class centers via Maximum Likelihood or Gradient Descent
- ◉ Assuming class priors as equivalent to each other or estimating them from the training data.
- ○ Assuming the class covariance matrices as full or diagonal
- ○ Assuming the class means as same or not

## Q7
1 Point

For Kernel Regression, which one of the following factors affects the trade-off between overfitting and underfitting the most?

- ○ Type of the kernel function
- ○ Form of the distance metrics (L1, L2, etc.)
- ○ Maximum height of the kernel function
- ◉ Width of the kernel

## Q8
4 Points

Assuming that you want to learn a decision tree (without any pruning) using the categorical features $X_1$, ..., $X_m$ and the categorical output label $Y$ on a training data. Please state whether the following are true or false.

### Q8.1
1 Point

If $X_i$ and $Y$ are independently distributed variables, then decision tree ignores $X_i$.

- ◉ True
- ○ False

### Q8.2
1 Point

If $IG(Y|X_i) = 0$, then decision tree ignores $X_i$.

- ○ True
- ◉ False

### Q8.3
1 Point

The maximum depth of the decision tree must be less than $logm$.

- ○ True
- ◉ False

## Q8.4

1 Point

If the number of training samples is $R$, then the maximum depth of the decision tree must be less than $1 + log_2 R$

- ◉ True
- ○ False

## Q9

2 Points

Consider that you are given the following training data where each sample $x_i$ is represented via a scalar and each sample is associated with a real-valued output $y_i$:

| $x$ | $y$ |
|---|---|
| 0 | 2 |
| 2 | 2 |
| 3 | 1 |

Please answer the following questions accordingly.

## Q9.1

1 Point

When linear regression used, what will be the mean squared leave one out cross validation error? **Please write your answer as a fractional number considering the format a/b.**

## Q9.2
1 Point

When a model that always outputs a constant $y = c$ is learned from the non-left-out data points, , what will be the mean squared leave one out cross validation error? **Please write your answer as a fractional number considering the format a/b.**

## Q10
1 Point

Which one of the following statements about ensemble methods is true?

- (●) Combining weak learners via bagging reduces the variance, hence it is favorable.
- ( ) Combining strong learners using bagging reduces the variance, hence it is favorable.
- ( ) Combining strong learners via boosting reduces the bias, hence it is favorable.
- ( ) Combining weak learners using boosting reduces the variance, hence it is favorable.

## Q11
1 Point

Which one of the following statements about K-means is true?

- When $K = 1$, it gives the smallest value of the objective function
- For a predefined number of clusters, it minimizes the within class variance
- When the initial cluster centers are chosen from the training samples, it always converges to the global optimum.
- None of the above

## Q12
1 Point

Consider that $n$ denotes the number of training samples. Which one of the followings represent the running time of the nearest neighbor classification method?

- $O(1)$
- $O(logn)$
- $O(n)$
- $O(n^2)$

## Q13
1 Point

Which one of the followings is a problem you might face with when you use sigmoid activation functions in a neural network?

- Sigmoid function is a convex function and hence it cannot solve non-convex problems
- Sigmoid function might give negative activations and this affects learning badly.
- Sigmoid function might lead to gradients close to zero and this affects learning badly.
- None of the above

## Q14

1 Point

Suppose that the training data is given in the form of an n-by-n affinity matrix. Which one of the following methods can be used for clustering?

A) Agglomerative clustering

B) K-means

C) Normalized Cuts

- ○ A
- ◉ B
- ○ C
- ○ A and B
- ○ A and C
- ○ B and C
- ○ A, B, and C

## Q15

1 Point

Consider that the training data for a binary classification problem includes the following one-dimension features:

| $x$ | $y$ |
|---|---|
| 5 | 1 |
| 2 | 1 |
| -2 | -1 |

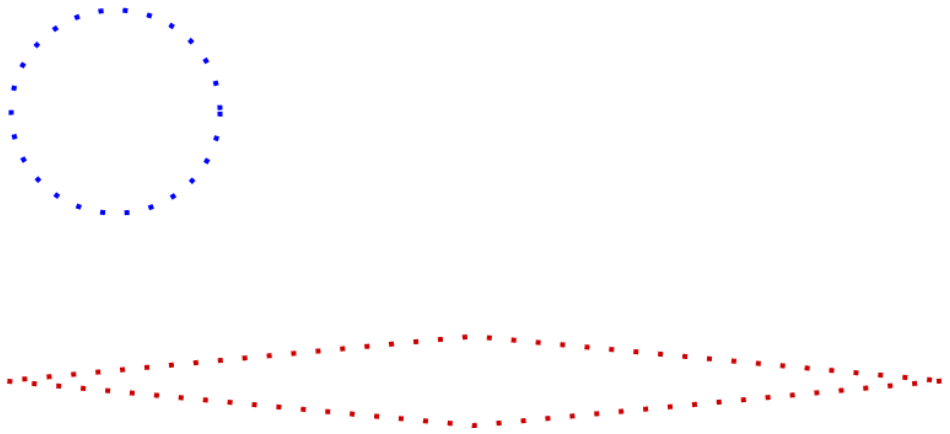
Which one of the followings represent the values for the model parameters ($w$ and $b$) given by a hard-margin SVM?

- [x] $w = 1, b = 1$
- [ ] $w = 0, b = 1$
- [ ] $w = 1, b = 0$
- [ ] $w = \infty, b = 0$

## Q16

1 Point

Suppose you want to cluster the data samples (dots) shown below into two clusters as illustrated in red and blue colors. Which one of the following clustering algorithms you prefer for this desired clustering result? **You can assume that Euclidean distance is used in each one of these algorithms.**



- [ ] Hierarchical clustering with complete link
- [ ] Hierarchical clustering with single link
- [x] Hierarchical clustering with average link
- [ ] None of the above

## Q17

1 Point

Which of the following classifiers are capable of achieving 100% training accuracy for the XOR problem? **Please mark all that apply. To get full credits, the set of all correct answers must be checked.**

- ○ AdaBoost with depth-two decision trees
- ○ Logistic regression
- ○ An artificial neural network with one hidden layer
- ◉ AdaBoost with depth-one decision trees

- ☐ AdaBoost with depth-two decision trees
- ☐ Logistic regression
- ☑ An artificial neural network with one hidden layer
- ☑ AdaBoost with depth-one decision trees

## Q18

1 Point

Which of the following statements are valid for support vector machines? **Please mark all that apply. To get full credits, the set of all correct answers must be checked.**

- ○ Increasing the value of $C$ is likely to decrease the training error
- ◉ Increasing the value of $C$ is likely to decrease the margin
- ○ Increasing the value of $C$ is likely to increase the robustness of the model against outliers
- ○ Hard-margin SVM and soft-margin SVM become equivalent when $C = 0$

- [ ] Increasing the value of $C$ is likely to decrease the training error
- [x] Increasing the value of $C$ is likely to decrease the margin
- [x] Increasing the value of $C$ is likely to increase the robustness of the model against outliers
- [ ] Hard-margin SVM and soft-margin SVM become equivalent when $C = 0$

## Q19

1 Point

Which of the following statements are valid for decision trees? **Please mark all that apply. To get full credits, the set of all correct answers must be checked.**

- There exists a split such that Information gain (IG) is positive if a given node includes samples from at least two classes
- The deeper the tree is, it is more likely that you are observing overfitting
- IG at each level does not increase as you go further down to a leaf node,
- Random forests are less likely to overfit than decision trees

- [x] There exists a split such that Information gain (IG) is positive if a given node includes samples from at least two classes
- [x] The deeper the tree is, it is more likely that you are observing overfitting
- [ ] IG at each level does not increase as you go further down to a leaf node,
- [x] Random forests are less likely to overfit than decision trees

## Q20
1 Point

Which of the following statements are valid for hierarchical clustering? **Please mark all that apply. To get full credits, the set of all correct answers must be checked.**

- ○ The number of clusters is a hyperparameter
- ○ The bottom-up agglomerative clustering algorithm repeatedly merges the two clusters that minimize the distance between clusters
- ○ Complete link works only with the Euclidean distance metric
- ○ During bottom-up agglomerative clustering, single link is more sensitive to outliers than complete link

- ☑ The number of clusters is a hyperparameter
- ☑ The bottom-up agglomerative clustering algorithm repeatedly merges the two clusters that minimize the distance between clusters
- ☑ Complete link works only with the Euclidean distance metric
- ☐ During bottom-up agglomerative clustering, single link is more sensitive to outliers than complete link

## Q21
1 Point

Which of the following statements are valid for spectral clustering? **Please mark all that apply. To get full credits, the set of all correct answers must be checked.**

- The Fiedler vector is the eigenvector associated with the second largest eigenvalue of the Laplacian matrix
- Finding the optimal Ncut in polynomial time is not possible
- The value of the scale parameter has a direct effect on the clustering results
- The Laplacian matrix of a graph is invertible

- [ ] The Fiedler vector is the eigenvector associated with the second largest eigenvalue of the Laplacian matrix
- [ ] Finding the optimal Ncut in polynomial time is not possible
- [x] The value of the scale parameter has a direct effect on the clustering results
- [x] The Laplacian matrix of a graph is invertible

## Q22

1 Point

Which of the following statements are valid for the k-NN algorithm? **Please mark all that apply. To get full credits, the set of all correct answers must be checked.**

- You can use k-NN for both classification and regression.
- As k increases, the bias usually increases.
- The smaller the value of k is, the smoother the decision boundary.
- As k increases, the variance usually increases.

- [x] You can use k-NN for both classification and regression.
- [x] As k increases, the bias usually increases.
- [x] The smaller the value of k is, the smoother the decision boundary.
- [ ] As k increases, the variance usually increases.

# Final Exam - Part 1

GRADED

STUDENT
MEHMET TAHA USTA

TOTAL POINTS
**17 / 44 pts**

QUESTION 1

Exam regulations **0** / 0 pts

QUESTION 2

(no title) **3** / 3 pts

2.1 (no title) **1** / 1 pt

2.2 (no title) **1** / 1 pt

2.3 (no title) **1** / 1 pt

QUESTION 3

(no title) **10** / 18 pts

3.1 (no title) **0** / 1 pt

3.2 (no title) **0** / 1 pt

3.3 (no title) **0** / 1 pt

3.4 (no title) **1** / 1 pt

3.5 (no title) **0** / 1 pt

| | | |
|---|---|---|
| 3.6 | (no title) | **1** / 1 pt |
| 3.7 | (no title) | **1** / 1 pt |
| 3.8 | (no title) | **1** / 1 pt |
| 3.9 | (no title) | **1** / 1 pt |
| 3.10 | (no title) | **0** / 1 pt |
| 3.11 | (no title) | **1** / 1 pt |
| 3.12 | (no title) | **0** / 1 pt |
| 3.13 | (no title) | **0** / 1 pt |
| 3.14 | (no title) | **1** / 1 pt |
| 3.15 | (no title) | **0** / 1 pt |
| 3.16 | (no title) | **1** / 1 pt |
| 3.17 | (no title) | **1** / 1 pt |
| 3.18 | (no title) | **1** / 1 pt |

QUESTION 4

(no title) **0** / 1 pt

QUESTION 5

(no title) **0** / 1 pt

QUESTION 6

(no title) **0** / 1 pt

QUESTION 7

(no title) **1** / 1 pt

QUESTION 8

(no title) **2** / 4 pts

| | | |
|---|---|---|
| 8.1 | (no title) | **0** / 1 pt |
| 8.2 | (no title) | **1** / 1 pt |
| 8.3 | (no title) | **1** / 1 pt |
| 8.4 | (no title) | **0** / 1 pt |

QUESTION 9

(no title) **0** / 2 pts

| | | |
|---|---|---|
| 9.1 | (no title) | **0** / 1 pt |
| 9.2 | (no title) | **0** / 1 pt |

QUESTION 10

(no title) **0** / 1 pt

QUESTION 11

(no title) **1** / 1 pt

QUESTION 12

(no title) **0** / 1 pt

QUESTION 13

(no title) **0** / 1 pt

QUESTION 14

(no title) **0** / 1 pt

QUESTION 15

(no title) **0** / 1 pt

QUESTION 16

(no title) **0** / 1 pt

QUESTION 17

(no title) **0** / 1 pt

QUESTION 18

(no title) **0** / 1 pt

QUESTION 19

(no title) **0** / 1 pt

QUESTION 20

(no title) **0** / 1 pt

QUESTION 21

(no title) **0** / 1 pt

QUESTION 22

(no title) **0** / 1 pt