Machine Learning I: Final Exam: 2022

ⓘ This is a preview of the published version of the quiz

Started: Mar 24 at 5:19pm

Quiz Instructions

Machine Learning I

Final Exam (Q3 2022_1)

Prof. Dr. Gregory Wheeler

March 25, 2022

INSTRUCTIONS:

The exam consists of **18 questions** of which you will have to answer all questions. The question format is **multiple choice**, **multiple answer**. You have **60 minutes** to complete the exam. The maximum score is **50 points**.

The exam is open book but closed device: you may use any books or printed notes you wish, but you may not use any electronic device (except the browser window used to take this exam).

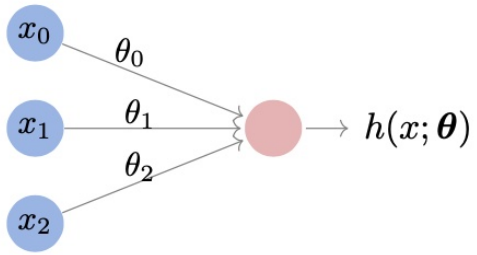
Good luck!

Question 1 — 3 pts

Suppose that x_1 and x_2 are binary variables that take either the value 0 or the value 1. Suppose the boolean function $x_1 \not\mapsto x_2$ is defined as follows:

x_1	x_2	$x_1 \not\mapsto x_2$
1	1	1
1	0	1
0	1	0
0	0	0

With respect to the following logistic unit,



what values for $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2]$ are such that $h(x;\boldsymbol{\theta})$ approximates $x_1 \not\mapsto x_2$?

- [ ] [10, -30, 10]
- [ ] [-20, 30, 10]
- [ ] [-20, 30, -10]
- [ ] None of the available answers.

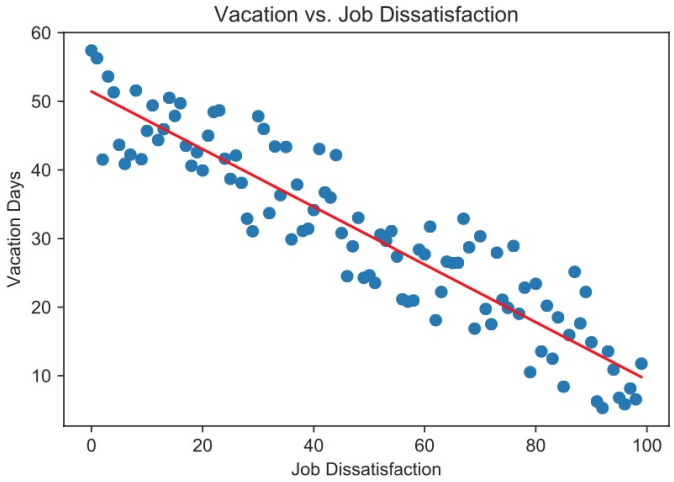
☐ [-10, 30, -10]

☐ [10, -20, 30]

Question 2

3 pts

The following scatter plots shows the relationship between days of vacation afford to an employee (y) and the level of employee dissatisfaction (x), measured from no dissatisfaction (0) to complete dissatisfaction (100).



☐ A) $\hat{y} = 51.43 + 0.42x$

☐ B) $\hat{y} = 51.43 - 0.42x$

☐ C) $\hat{y} = -51.43 + 0.42y$

☐ D) $\hat{x} = 51.43 + 0.42y$

☐ E) $\hat{x} = -0.42y + 51.43y$

☐ F) $\hat{x} = 0.42y + 51.43y$

- [ ] A
- [ ] D
- [ ] B
- [ ] F
- [ ] C
- [ ] E

Question 3

3 pts

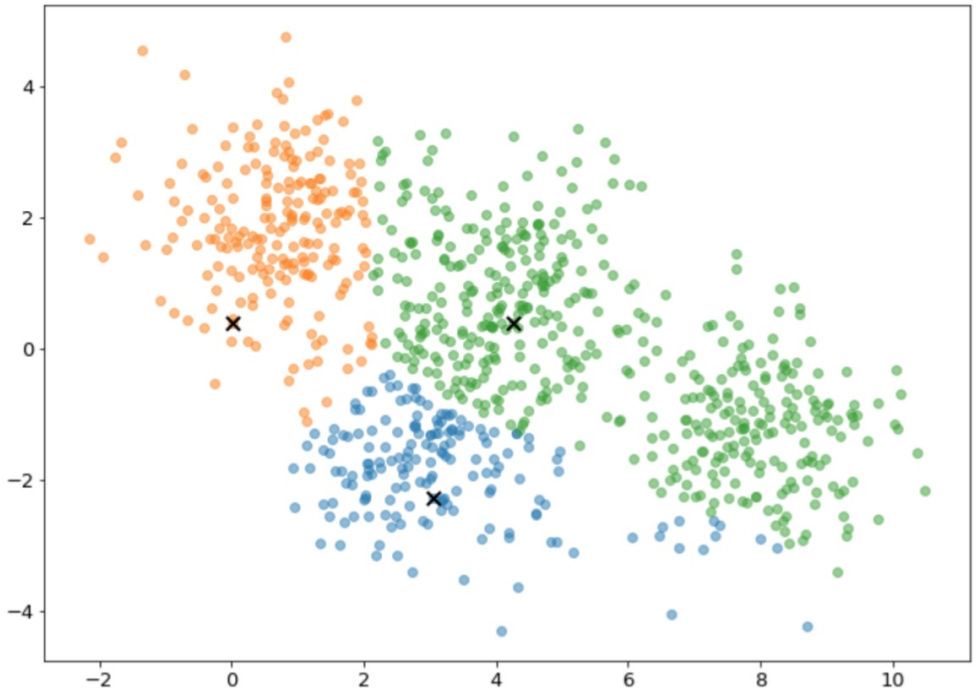
Ridge regression. Select all and only that apply:

- [ ] adds an L1-norm penalty to the cost function
- [ ] often results in setting several parameters to zero
- [ ] the regularization penalty lambda applies to all parameters
- [ ] None of the available answers.
- [ ] reduces variance at the expense of higher bias
- [ ] is more sensitive to outliers than ordinary least-squares (OLS) regression

Question 4

3 pts

The following plot is of unlabeled data points in $\mathbb{R}^2$.



Consider the following statements about the K-means algorithm in relation to the above plot. Select all and only that apply.

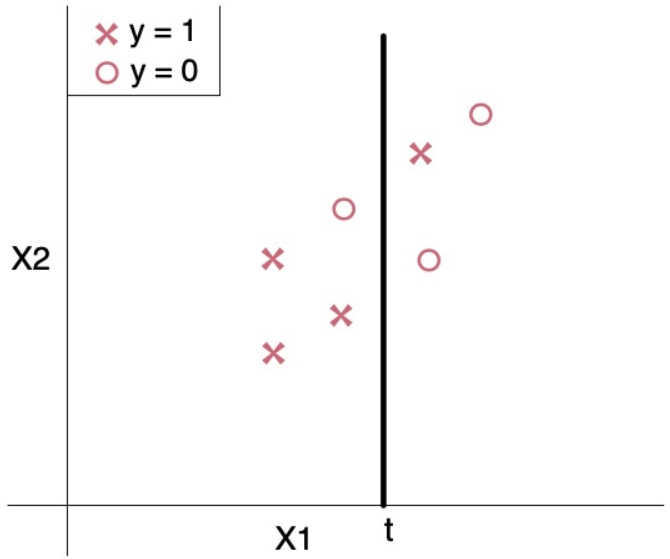
- [ ] K means has converged to local minima
- [ ] The plot depicts the result after executing the second loop of K-means

- [ ] K = 3
- [ ] K means has not converged to any local minima
- [ ] The plot depicts the result after executing the first inner-loop of K-means.
- [ ] None of the available answers.

Question 5

3 pts

Consider the following graph of labeled data in $\mathbb{R}^2$ and a single split of the data X1 < t.



Given this split, what is the probability that an observed point is in the positive class?

- [ ] 3/7
- [ ] 2/3
- [ ] 1/3
- [ ] 3/4
- [ ] 1/4
- [ ] 4/7

Question 6

3 pts

Suppose you are to select models from a set of candidate models according to the following two criteria:

(a) It must have a precision rate of at least 90%

(b) It must have a false negative rate of at most 10%

Below are vector encoded confusion matrices of your candidate models. That is, for the confusion matrix

		Actual Class	
		1	0
Predicted Class	1	w	x
	0	y	z

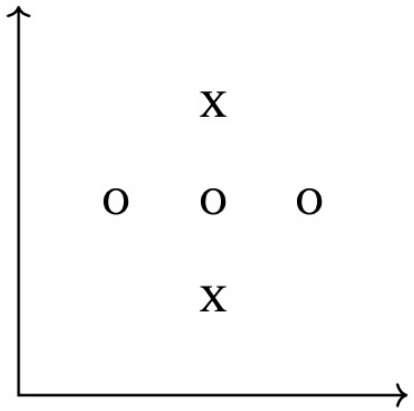
its vector encoding is $[w, x, y, z]$. Which (if any) satisfy both (a) and (b)?

- [ ] [90, 4, 10, 96]
- [ ] [82, 2, 18, 98]
- [ ] [79, 1, 21, 99]
- [ ] [78, 9, 22, 91]
- [ ] None of the available answers.
- [ ] [92, 20, 8, 80]

Question 7

3 pts

Suppose you have the following labeled training data $\mathcal{D}$ in $\mathbb{R}^2$:



Which of the following models, when trained appropriately, can produce zero classification errors on $\mathcal{D}$? Select all and only that are true.

- [ ] Decision Trees
- [ ] Principal Components Analysis (PCA)
- [ ] Linear Regression
- [ ] K-means Algorithm
- [ ] Neural Network with two layers (L = 2)
- [ ] Logistic Regression

Question 8

3 pts

The odds compare the probability of success (q) to the probability of failure ($1-q$). Suppose the thresholds probability of your classifier divides your $m = 9$ data points into the positive class ($y = 1$) and the negative class ($y = 0$) as follows:

$$\mathrm{PNPP}_{\wedge}\mathrm{PPNNN}$$

Which statements are then true? Select all and only that apply.

- [ ] According to your classifier, the odds that y = 1 is 2/5
- [ ] According to your classifier, the odds that y = 1 is 3.
- [ ] According to your classifier, the odds that y = 1 is 3/4
- [ ] According to your classifier, the odds that y = 0 is 2/5
- [ ] According to your classifier, the odds that y = 0 is 2/3
- [ ] According to your classifier, the odds that y = 0 is 1.5

Question 9

3 pts

Suppose your neural network's predictions in your training data are exactly right such that your training errors are exactly 0. In that circumstance, which of the following statements would be true? Mark all and only those that apply.

- [ ] The updates to your weights will be 0.
- [ ] The updates to your weights will depend on the activation functions
- [ ] The slope of your loss function with respect to your prediction is 0
- [ ] None of the available answers.
- [ ] The prediction errors in your test set will be 0.
- [ ] The updates to your weights will depend on the input data.

Question 10

3 pts

CART Models. Select all and only that apply

- [ ] Increasing maximum tree depth reduces variance.
- [ ] The depth of a learned decision tree cannot be larger than the number of training examples used to create the tree.
- [ ] Decision trees are robust to small changes to input data.
- [ ] The purity measures Gini and Entropy are different ways to measure the accuracy of a CART classification model on test data.
- [ ] Random Forests are a bootstrap aggregating (Bagging) method to reduce variance.

☐ Increasing maximum tree depth reduces bias

Question 11

3 pts

Consider the following model description.

```
RandomForestClassifier(max_depth = 5, criterion = 'gini', n_estimators=200)
```

Select all and only those that apply.

☐ A) There are 200 decision trees in this random forest.

☐ B) Each estimator in the random forest has at most 200 leaf nodes.

☐ C) The quality of a split is measured by $-\sum_{c=1}^{C} \hat{\pi}_c \log \hat{\pi}_c$, where $1 - \hat{\pi}_c$ is the misclassification rate for class c.

☐ D) The quality of a split is measured by $\sum_{c=1}^{C} \hat{\pi}_c (1 - \hat{\pi}_c)$, where $1 - \hat{\pi}_c$ is the misclassification rate for class c.

☐ E) Each estimator has at most 5 leaf nodes.

☐ F) None of the above

☐ B

☐ A

☐ E

☐ F

☐ C

- [ ] D

Question 12

3 pts

Suppose you fit an OLS regression for two data sets:

(a) a UN dataset estimating average life expectancy of each country's citizens from economic indicators for each country.

(b) a synthetic dataset constructed by taking random samples from a normal distribution and assigning to each sample a value *y* that is a linear function *f* of x and noise ϵ, also drawn from a known normal distribution, that is: $y = f(x) + \epsilon$.

Consider the bias-variance trade-off for (a) and (b). Which of the following, if any, can we estimate with high accuracy?

- [ ] The bias component of MSE for dataset (a)
- [ ] the bias component of MSE for dataset (b)
- [ ] the variance component of MSE for dataset (a)
- [ ] the variance component of MSE for dataset (b)
- [ ] None of the available answers.

Question 13

3 pts

Which statements about Dropout Regularization are true? Select all and only that apply.

- [ ] Dropout increases the capacity of a neural network.
- [ ] Dropout can be applied to all layers of a neural network.
- [ ] Dropout helps to reduce feature co-adaptation.
- [ ] Dropout introduces sparsity in the hidden units of a network
- [ ] None of the available answers.
- [ ] Dropout addresses underfitting.

Question 14

3 pts

Select all and only that are true statements about back propagation.

- [ ] It sequentially calculates and stores intermediate variables defined by the neural network architecture.
- [ ] Weights are gradually changed to minimize error over the training data.
- [ ] It sequentially calculates and stores the gradients of intermediate variables and parameters within a neural network.
- [ ] Nodes are added or removed randomly to improve generalization.
- [ ] Weights are replaced to match one training example at a time.

- [ ] None of the available answers.

Question 15

2 pts

Consider the following lines of Python code.

```
A = np.array([[17, 30, 25, 1],
              [48, 74, 12, 0],
              [97, 33, 53, 1]])
A[1,2]
```

After execution, what is the output? Select all and only that apply.

- [ ] 25
- [ ] TypeError
- [ ] 12
- [ ] 97
- [ ] 74
- [ ] 33

Question 16

2 pts

Consider the following lines of Python code.

```
x = 1
y = 2
x += x * y + 1
```

After execution, what is the value of `x`? Select all and only that apply.

- [ ] 4
- [ ] False
- [ ] True
- [ ] 2
- [ ] 1
- [ ] 3

Question 17

2 pts

Recall the bias-variance trade-off. Which of the following is more harmful to the test error than the training error? Select all and only that apply.

- [ ] Expectation
- [ ] Variance
- [ ] Loss
- [ ] Risk
- [ ] Bias
- [ ] Utility

Question 18

2 pts

Suppose you construct a logistic regression model and the odds of an observation *x* belonging to the positive class is 3. What proportion of *x's* would you expect to belong to the positive class?

- [ ] 66%
- [ ] 33%
- [ ] 60%
- [ ] 67%
- [ ] None of the available answers.
- [ ] 75%

In case you are facing problems with viewing tables, graphs, please open the **PDF version** (https://frankfurtschool.instructure.com/courses/6797/files/404715/download?download_frd=1) of the exam.

Quiz saved at 5:19pm Submit Quiz