Quiz 1 - Machine Learning Concepts

Due Feb 6 at 2:30pm **Points** 10 **Questions** 5 **Time Limit** 10 Minutes
Allowed Attempts 2

Attempt History

	Attempt	Time	Score
KEPT	**Attempt 2**	2 minutes	10 out of 10
LATEST	**Attempt 2**	2 minutes	10 out of 10
	Attempt 1	7 minutes	10 out of 10

Score for this attempt: **10** out of 10
Submitted Mar 5 at 2:49pm
This attempt took 2 minutes.

Question 1

2 / 2 pts

Which of the following statement about machine learning is correct?

- The learning mainly relies on the theoretical analysis of the relationship between labels and features.
- Correct! Clustering is one of the most-often used unsupervised learning methods.
- Some machine learning methods do not require data.
- Regression and classification are unsupervised learning.

Question 2

2 / 2 pts

The major difference between supervised learning and unsupervised learning is

Correct!

- (selected) The correct answers are included in the data for supervised learning.
- There is no training in unsupervised learning.
- Unsupervised learning can perform better with smaller dataset.
- There is no mathematical models in unsupervised learning.

Question 3

2 / 2 pts

What is a model in supervised learning?

Correct!

- (selected) A mathematical mapping between labels and features.
- A representative subset of the data
- The correct answers to the prediction problem
- Theoretical derivation using domain knowledge

Question 4

2 / 2 pts

What terms do we use to refer to the correct answers in a dataset?

- response
- dependent variable

Correct!

- All of the others
- target variable
- labels

Question 5

2 / 2 pts

You want build a machine learning model to predict whether a customer would choose a product on your website. Which of the following statements is correct?

Correct!

- All of the others
- The labels should be whether a customer selects a product.
- The customer rating of a product can be a good feature.
- The dataset should contain the information of many customers and their product choice.
- The product price can be a good feature.

Quiz Score: **10** out of 10

Quiz 2 - Linear Regression

Due Feb 13 at 4:10pm **Points** 10 **Questions** 5 **Time Limit** 10 Minutes
Allowed Attempts 2

Instructions

You have 10 minutes and two attempts for this quiz.

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	6 minutes	10 out of 10

Score for this attempt: **10** out of 10
Submitted Feb 13 at 3:48pm
This attempt took 6 minutes.

Question 1 — 2 / 2 pts

Which statement about p-value in a linear regression is true?

- The larger the p-value, the more the dependent variable increases when the predictor increases.
- The smaller the p-value, the higher the predicted value of the dependent variable
- The smaller the p-value, the more significant the predictor variable is

Correct!

The larger the p-value, the more impactful the predictor variable is.

Question 2

2 / 2 pts

Which type of graphs is suitable to explore the relationship between two interval variables?

Bar

Correct!

Scatterplot

Histogram

Pie

Question 3

2 / 2 pts

A variable measures the education level of a customer as follows: under high school, high school, college, and post-graduate. How many dummy variables do we need to add this variable to a linear regression?

1

4

Correct!

3

2

Question 4

2 / 2 pts

R-squared indicates the significance of all independent variables.

- ○ True
- ◉ False

Correct!

Question 5

2 / 2 pts

The linear regression model can only captures linear relationships between dependent and independent variables.

- ○ True
- ◉ False

Correct!

Quiz Score: **10** out of 10

Quiz 3 - Improvements for Linear Models

Due Feb 25 at 11:59pm **Points** 10 **Questions** 5
Time Limit 15 Minutes **Allowed Attempts** 2

Instructions

You have two attempts.

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	3 minutes	10 out of 10

Score for this attempt: **10** out of 10
Submitted Feb 25 at 1:42am
This attempt took 3 minutes.

Question 1 — 2 / 2 pts

Which of the following should be considered as a categorical variable?

- Education measured in the number of years in school
- House price in dollars
- House type: condo - 1, townhouse - 2, single house - 3 (Correct!)
- House size in sqft

Question 2

2 / 2 pts

Which of the following code change the data type of a variable?

- class(cars$doors)
- doors <- names(cars$doors)
- as.factor(cars$doors) — Correct!
- unique(cars$doors)

Question 3

2 / 2 pts

Which type of graph explores the relationship between an interval variable and a categorical variable?

- histograms of the categorical variable
- Boxplots of the interval variable across categories — Correct!
- Separated scatterplots of the interval variable
- Bar charts across categories

Question 4

2 / 2 pts

Which of the following statement is correct?

There should be three coefficients estimated for a categorical variable with three levels.

The base case can only be the one with the lowest value

Correct!

The interpretation of the coefficients of a categorical variable are always in reference to the base case.

The omitted case of a categorical variable suggests it has no impact on the dependent variable.

Question 5

2 / 2 pts

We want to predict a country's happiness index using economic and social features of a country such as GDP, freedom index, healthcare conditions, etc. We expect that the impact of freedom index might be moderated by the richness of a country. That means, people in poor economic conditions may care less about freedom for their happiness than people in richer countries. What model specification could help us to incorporate such consideration?

Correct!

An interaction term between GDP and freedom index.

A square term of freedom index

Change GDP into a categorical variable as poor vs rich countries

Delete GDP from the model

Quiz Score: **10** out of 10

Quiz 4 - Logistic Regression

Due Mar 5 at 11:59pm **Points** 10 **Questions** 5 **Time Limit** 15 Minutes
Allowed Attempts 2

Attempt History

	Attempt	Time	Score
KEPT	**Attempt 2**	4 minutes	10 out of 10
LATEST	**Attempt 2**	4 minutes	10 out of 10
	Attempt 1	6 minutes	8 out of 10

Score for this attempt: **10** out of 10
Submitted Mar 5 at 2:42pm
This attempt took 4 minutes.

Question 1

2 / 2 pts

The R function we use to estimate a logistic regression model is lm().

- True
- False

Correct!

Question 2

2 / 2 pts

Which of the following statement is correct?

The probability is a linear function of independent variables

The odds is always between 0 and 1

Correct!

The odds of probability 0.5 is 1

The logit is always between 0 and 1

Question 3

2 / 2 pts

Which of the following statement is NOT correct?

The target variable of logistic regression is categorical variable.

Correct!

Logistic regression models the absolute value of the target variable.

Logistic regression is a method of classification.

Logistic regression models a curve between 0 and 1.

Question 4

2 / 2 pts

Which of the following statement is correct?

The odds has a linear relationship with independent variables.

Correct!

The logit has a linear relationship with independent variables.

The logistic regression we have learned can be directly applied to a variable with more than two categories.

The probability has a linear relationship with independent variables.

Question 5

2 / 2 pts

How is the coefficient (beta) of an interval variable X in a logistic regression associated with the predicted values?

Correct!

With one unit increase in X, the odds ratio of after vs. before is e^(beta).

With one unit increase in X, the probability increases by e^(beta) units.

With one unit increase in X, the odds increase by beta units.

With one unit increase in X, the probability increases by beta units.

Quiz Score: **10** out of 10

Quiz 5 - Generative Models

Due Mar 5 at 11:59pm **Points** 10 **Questions** 6 **Time Limit** 15 Minutes
Allowed Attempts 2

Attempt History

	Attempt	Time	Score
KEPT	Attempt 2	1 minute	10 out of 10
LATEST	Attempt 2	1 minute	10 out of 10
	Attempt 1	5 minutes	10 out of 10

Score for this attempt: **10** out of 10
Submitted Mar 5 at 2:53pm
This attempt took 1 minute.

Question 1

2 / 2 pts

Which of the following statement is correct about LDA, QDA and Naïve Bayes methods?

- They can be applied to classification of more than two categories.
- They are generative methods.
- All of the above. (Correct!)
- They are based on Bayesian Theorem.

Question 2

2 / 2 pts

Which of the following probability are we trying to estimate in discriminant analysis models?

- p(X = x)
- p(Y = y)

Correct!

- p(Y = y | X = x)
- p(X = x | Y = y)

Question 3

2 / 2 pts

Which of the following situation is LDA best suitable for?

- Large sample; the distribution of Xs does not matter.
- Small sample; the distribution of Xs is approximately normal only within one of the classes.

Correct!

- Small sample; the distribution of Xs within each class is approximately normal
- Large sample; the distribution of Xs is not normal.

Question 4

2 / 2 pts

Which of the following is Naïve Bayes best suitable for?

Correct!

- (selected) No distribution assumption but Xs are independent within each class.
- Normal distribution of Xs within each class
- No assumption about Xs
- No assumption about Xs distribution or independence

Question 5

1 / 1 pts

Which of the following is sensitivity?

- The ratio between predicted negatives and all observations
- The ratio between predicted positives and all observations

Correct!

- (selected) The ratio between predicted positives and all True positives
- The ratio between predicted negatives and all True negatives

Question 6

1 / 1 pts

Which of the following describes a Type II error?

-

Correct!

- (selected) A person was not diagnosed with cancer, despite having cancer.

- A healthy person is diagnosed with cancer.

Quiz Score: **10** out of 10