Exam I

Due No due date **Points** 10 **Questions** 50
Available Oct 10 at 11:30am - Oct 10 at 1pm about 2 hours **Time Limit** 75 Minutes

Instructions

This exam is 50 questions and you only have 75 minutes to complete so please pace appropriately.

Also, questions are shown sequentially one at a time, and your answers are locked after you submit each question, so be sure of your answer to a question before you move on to the next one.

This quiz was locked Oct 10 at 1pm.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	29 minutes	9.5 out of 10

Score for this quiz: **9.5** out of 10
Submitted Oct 10 at 12:01pm
This attempt took 29 minutes.

Question 1

0.2 / 0.2 pts

Select all scenarios that are examples of supervised learning

Correct!

- [x] Predicting a buyer's chance of clicking on an online advertisement based on the previous behavior of similar online shoppers.
- [ ] Using the nucleotide sequences on a region of non-coding DNA shared among species to estimate a phylogenetic tree.

An infant, unable to speak, but forming concepts of ‘r’ or ‘l’ sounds based on the grouping of similar sounds over time.

Correct!

Netflix using their database of user ratings to predict how you would rate a movie you haven’t seen

Question 2

0.2 / 0.2 pts

Select all examples of semi-supervised learning (as opposed to pure supervised or unsupervised learning examples)

Making stock predictions for a high-frequency trading company

Determining a taxonomy (tree-like classification) for animals based solely on their features

Correct!

Collecting constant GPS data, automatically clustering repeated locations, then having a personal label those clusters as “home” or “work” with the goal of having the setup detect whenever the wearer is at home or at work.

Correct!

Your learning of music genres, especially your ability to ask questions about a category of music that you notice as particularly distinct.

Question 3

0.2 / 0.2 pts

Weather forecasters in Denton decided to build a model that predicts tomorrow’s high temperature from the previous 30 day’s high temperatures.

To do this, they used the past year's weather data to train the model. They had perfect accuracy in predicting when using last year's data for testing. However, when they applied the same model to predict the weather the next day, they found it was off by 10 degrees.

Select all statements that are likely to apply to their model.

Correct!

- [x] They should have used separate sets of data for training and for testing to pick the right model
- [ ] It's Denton. It's not possible to predict the weather. They should just give up.

Correct!

- [x] The model is overfitting (it is too complex, too many variables)
- [ ] The model is overgeneralizing (it is too simple)

Question 4

0.2 / 0.2 pts

Select all below that are an example of model overfitting (instead of overgeneralization)

- [ ] You have an idea. You are often correct with your ideas. therefore since you came up with the idea, it must be right.
- [ ] Fitting a U-shaped curve with a straight line

Correct!

- [x] This often happens when your model has too many parameters for too
 little data

Correct!

- [x] Fitting a U-shaped curve with a 20 degree polynomial, that fits all 21
 points perfectly.

Question 5

0.2 / 0.2 pts

If k-nearest neighbors was your model of how you make decisions, which value of k would be more likely to be superstitious (lead to poor generalization, fit to the noise, be “too complex” of a model…)

- k=100
- Correct! k=1
- k as it approaches infinity
- k=5

Question 6

0.2 / 0.2 pts

If you were trying to create a system to predict whether a picture is of a city or country scene, match the corresponding concept in machine learning terminology.

Correct! **Sample** — a picture

Correct! **Feature** — the percentage of blue pix

Correct! **Target** — a label of 'City' or 'Country

'ou Answered

Correct Answer

Question 7

0.2 / 0.2 pts

Select all the ways you may be able to improve the accuracy of your machine learning prediction.

Correct! ☑ Pick a different learning model (kNN, SVM, ...)

Correct! ☑ Derive/predict new features from current features in your data set

Correct! ☑ Add/remove features

Correct! ☑ Change your hyperparameter (e.g. k in kNN)

Correct! ☑ Get more samples

Question 8

0.2 / 0.2 pts

What is the best rule of thumb for the number of samples needed for good machine model learning.

☐ features < samples

Correct! ☑ samples > feature^2

☐ samples > features

☐ features < samples^2

Question 9

0.2 / 0.2 pts

Match the advanced machine learning concept to an example of it

Correct! **Clustering** — babies learning that 'r' and

Correct! **Feature engineering** — creating 'day of the week'

Correct! **Feature selection** — Including "IQ" and "GPA" ir

Correct! **Deep learning** — a type of machine learning

Question 10

0.2 / 0.2 pts

If you are picking among many different model variants you generally split your data into three different groups. Match the group to a property of that group.

Correct! **Training set** — data used to create the mc

Correct! **Validation set** — like a test set but it is used

Correct! **Test set** — the data set used to evalua

Correct!

Question 11

0.2 / 0.2 pts

The difference between regression and classification is that regression predicts nominal variables/labels, while classification predicts continuous variables.

- True
- False

Correct!

Question 12

0.2 / 0.2 pts

Regression models can only use continuous variables as features and classification models can only use categorical/nominal variables as features.

- True
- False

Correct!

Question 13

0.2 / 0.2 pts

Match the method of normalization with the result

Correct!

new_score = (old_score - min) / (max - min) — range between 0 and 1

Correct!

new_score = (old_score - mean) / standard_deviation — 0 mean and unit variance

Question 14

0.2 / 0.2 pts

Let's assume you expect your friend carrier a double-headed coin in their pocket 0.1% of the time. Match the following probabilities. Hint: you don't need to calculate them explicitly (use your intuitions), but you can if you want to.

Correct!	**P(10 heads|a trick coin)**	1
Correct!	**p(a fair coin)**	0.999
Correct!	**p(10 heads in a row | a fair coin)**	1/2^10 = ~0.001
Correct!	**p(a trick coin | 10 heads in a row)**	~0.2
Correct!	**p(10 heads in a row)**	~ 0.002

Question 15

0.2 / 0.2 pts

Match the terminology with the appropriate conditional probability

Correct!	**p(data | model)**	likelihood
Correct!	**p(model | data)**	posterior (the most sough
Correct!	**p(model)**	prior

Correct!

p(data) normalization term

Question 16

0.2 / 0.2 pts

Which are true about Naive Bayes classifiers?

Correct!

The most common functional form used in naive bayes is gaussian (hence "Gaussian Naive Bayes or GNB) since all that is needed is the mean and variance of the data for each class to calculate the likelihood.

Correct!

Although Bayes rule is a mathematical truth, the assumption of independent features and gaussian probability functions are just approximations that, though efficient, can lead to misclassifications

Correct!

Naive Bayes classification is fast since each feature can be treated independently and the functional form for the probabilities are often simple to calculate.

Correct!

The "Naive" part is because we assume the likelihoods of each feature are independent (so the overall probability is found by simply multiplying them)

Question 17

0.2 / 0.2 pts

When using the k-Nearest Neighbors classifier, the k represents the number of nearest neighbors from the training set to find, and the class is the most represented class among those neighbors.

Correct!

- (●) True
- ( ) False

Question 18

0.2 / 0.2 pts

Normalization is not important in k-NN classification because the features with the larger range should always have a larger influence on the k-NN distance metric than other features.

- ( ) True

Correct!

- (●) False

Question 19

0.1 / 0.2 pts

Which equations are linear equations if the variables are x's and constants are a's?

orrect Answer

- [ ] $y = a_0 x_0 + x_1 + a_2 x_2$
- [ ] $y = a_0 x_0 + a_1 x_1 + a_2 (x_2)^2$

Correct!

- [x] $y = a_0 + a_1 x_1 + a_2 x_2$
- [ ] $y = a_0 x_0 x_1 + a_2 x_2$

Question 20

0.2 / 0.2 pts

Match the goal for a linear regression model with the appropriate introduced feature

Correct!	**have an intercept (e.g. the "b" in "y = a x + b")**	Introduce a new feature w ▾
Correct!	**Create a quadratic (2nd degree polynomial fit) from a linear regression model**	add a feature that is the sc ▾
Correct!	**Use a regression model for binary classification**	change the target variable ▾

Question 21

0.2 / 0.2 pts

Match the regularization strategy with the expected outcome

Correct!	**No regularization**	Irrelevant features will no ▾
Correct!	**Ridge regression**	Irrelevant feature coefficie ▾
Correct!	**Lasso regression**	Irrelevant feature coefficie ▾

Question 22

0.2 / 0.2 pts

Which technique is more aptly considered a classifier?

- Ridge regression

Regularized regression

Lasso regression

Correct!

Logistic regression

Question 23

0.2 / 0.2 pts

What is the purpose of regularization in linear regression?

Correct!

To decrease the coefficient values for irrelevant terms in the regression model

Correct!

To improve prediction accuracy on a future test set better than ordinary linear regression

Correct!

To diminish the contribution of irrelevant features to the resulting model, effectively performing automated feature selection during learning

To lasso the ridge with an elastic net

Question 24

0.2 / 0.2 pts

Match the term with the definition using the following values in a classification task
TP: True positives
FP: False positives - classified a positive, but truly negative
TN: True negatives
FN: False negatives - classified as negative, but truly positive

Correct!

	Sensitivity (Recall for the positive case)	TP / (TP + FN)
Correct!	**Precision (for the positive case)**	TP / (TP + FP)
Correct!	**Accuracy**	(TP + TN) / (TP + TN + FP +
Correct!	**Specificity (Recall for the negative case)**	TN / (TN + FP)
Correct!	**F1 Score**	The harmonic mean of pre

Question 25

0.2 / 0.2 pts

Match the cross-validation technique with the description

Correct!	**K-fold cross-validation**	Train your model on K-1 gı
Correct!	**Leave one out cross-validation**	Same as K-fold cross-valid
Correct!	**Subject-wise cross-validation**	When you use the data frc
Correct!	**Nested cross-validation**	A cross-validation techniq

Question 26

0.2 / 0.2 pts

Match the regression error metric with the appropriate statement

Correct!

Root Mean Square Logarithmic Error — An error of 110 instead of

Correct!

Mean Square Error — An error of 110 instead of

Correct!

Mean Absolute Error — An error of 110 instead of

Question 27

0.2 / 0.2 pts

When you want to know how well a product will work on a new person without any individual-specific training, it is better to use K-fold cross-validation than subject-wise cross-validation, because K-fold cross-validation may have an individual's data in both the training and test sets, which is what you want in that case.

- True

Correct!

- (•) False

Question 28

0.2 / 0.2 pts

When you use cross-validation to select the right hyperparameter, you need a separate set of data to properly measure the accuracy of the model with that hyperparameter (due to potential overfitting). Unfortunately, many scientists don't do this (although they should)

Correct!

- (•) True

False

Question 29

0.2 / 0.2 pts

Match the term with the definition using the following values in a classification task
TP: True positives
FP: False positives - classified a positive, but truly negative
TN: True negatives
FN: False negatives - classified as negative, but truly positive

Correct!	**Sensitivity (Recall for the positive case)**	TP / (TP + FN)
Correct!	**Precision (for the positive case)**	TP / (TP + FP)
Correct!	**Accuracy**	(TP + TN) / (TP + TN + FP +
Correct!	**Specificity (Recall for the negative case)**	TN / (TN + FP)
Correct!	**F1 Score**	The harmonic mean of pre

Question 30

0.2 / 0.2 pts

Match the cross-validation technique with the description

Correct!	**K-fold cross-validation**	Train your model on K-1 g

Correct! **Leave one out cross-validation** — Same as K-fold cross-valid

Correct! **Subject-wise cross-validation** — When you use the data frc

Correct! **Nested cross-validation** — A cross-validation techniq

Question 31

0.2 / 0.2 pts

Match the regression error metric with the appropriate statement

Correct! **Root Mean Square Logarithmic Error** — An error of 110 instead of

Correct! **Mean Square Error** — An error of 110 instead of

Correct! **Mean Absolute Error** — An error of 110 instead of

Question 32

0.2 / 0.2 pts

When you want to know how well a product will work on a new person without any individual-specific training, it is better to use K-fold cross-validation than subject-wise cross-validation, because K-fold cross-validation may have an individual's data in both the training and test sets, which is what you want in that case.

- True
- Correct! False

Question 33

0.2 / 0.2 pts

When you use cross-validation to select the right hyperparameter, you need a separate set of data to properly measure the accuracy of the model with that hyperparameter (due to potential overfitting). Unfortunately, many scientists don't do this (although they should)

Correct!

- (•) True
- ( ) False

Question 34

0.2 / 0.2 pts

Match the concepts to the descriptions for support vector machines

Correct!

Support Vectors — The samples that are near ▾

Correct!

Margin — The distance between clas ▾

Correct!

Kernel — The functions used to proj ▾

Correct!

Slack Variable — A parameter to allow samp ▾

Question 35

0.2 / 0.2 pts

Check all of the features of the Random Forest classifier

Correct! ☑ Averages the result of a set of decision trees to arrive at a final prediction

Correct! ☑ Is an ensemble method using decision trees

☐ Relies on a "random walk" strategy

Correct! ☑ Creates trees using a subset of features

Question 36

0.2 / 0.2 pts

Of the following classifiers, select the three that are more likely to work well on a complex learning problem

Correct! ☑ Multi-layer neural network

☐ Perceptron

Correct! ☑ SVM with radial basis function kernel

☐ SVM with no kernel

Correct! ☑ Random Forest

☐ Decision tree

Question 37

0.2 / 0.2 pts

Match the following classifiers with the types of boundaries between classes they can construct in the feature space

Correct! **Decision trees** Vertical and Horizontal bo ▾

Correct! **Perceptron** A single separating line

Correct! **SVM with kernel functions, multilayer neural networks** Arbitrarily complex separa

Question 38

0.2 / 0.2 pts

The perceptron is simple, one-layer neural networks which is guaranteed to find a classification if the data is linearly separable. However, it fails on problems as simple as training for the XOR (exclusive-or) function.

Correct!

- (•) True
- ( ) False

Question 39

0.2 / 0.2 pts

Backpropagation is the ability of deep learning neural networks, like google's Deep Dream, to propagate signals from higher level neurons down to low-level neurons - allowing us in some cases to see the "hallucinated" images.

- ( ) True

Correct!

- (•) False

Question 40

0.2 / 0.2 pts

A random forest classifier uses boosting on a set of decision trees to increase performance

- True

Correct!

- False

Question 41

0.2 / 0.2 pts

Boosting combines the predictions of all the learners with equal weight.

- True

Correct!

- False

Question 42

0.2 / 0.2 pts

Bootstrap estimation is a statistical technique which relies on resampling (often with replacement) from a distribution of data to arrive at a distribution of parameter values for the data (e.g. means, variances...) rather than simply one number.

Correct!

- True
- False

Question 43

0.2 / 0.2 pts

In bagging, both the features (commonly the columns in a data set) and samples/observations (commonly the rows in a data set) may be sampled. And this can be done with or without replacement.

Correct!

- (●) True
- ( ) False

Question 44

0.2 / 0.2 pts

If you are teams #2 and #3 in a competition and you want to merge your teams to beat #1, you will more likely be using bagging rather than boosting to win.

- ( ) True

Correct!

- (●) False

Question 45

0.2 / 0.2 pts

Bagging is applied to learners that are more likely to overfit, while boosting is often applied to weak learners (to avoid overgeneralization). e.g. Bagging decision trees may have no limits on size, whereas boosted decision trees may be limited by depth.

Correct!

- (●) True
- ( ) False

Question 46

0.2 / 0.2 pts

The key characteristic distinguishing ensemble methods from general model selection strategies is that you retain the candidate models in ensemble methods and use all of them to arrive at a prediction.

Correct!

- (●) True
- ( ) False

Question 47

0 / 0.2 pts

Here is an analogy:

"Rose" is to "Flower" as "Porsche" is to "Automobile", because the first word is a type of the second word.

"North" is to "South" as "Black" is to "White" because second word is the opposite of the first word.

and so on...

The following is analogy can be said for four important concepts in machine learning. Fill in the blank.

Classification is to regression in supervised learning as _____ is to dimensionality reduction in unsupervised learning.

Or more succinctly

Classification is to regression as _____ is to dimensionality reduction

- ( ) Reinforcement learning
- (●) PCA
- ( ) Factor Analysis
- ( ) Clustering

'ou Answered

orrect Answer

Question 48

0 / 0.2 pts

Select all the model hyperparameters where a larger value leads to overfitting/higher model complexity rather than overgeneralization/simpler models.

Correct!

- [x] the degree of the polynomial in polynomial regression
- [ ] the k in k nearest neighbors

'ou Answered

- [x] the slack variable in support vector machine (larger means more slack or acceptance of errors))

orrect Answer

- [ ] the maximum depth parameter for decision trees

Question 49

0.2 / 0.2 pts

Which of the following is just an ensemble method applied to a simpler classifier?

Correct!

- [x] Random Forest
- [ ] Support Vector Machines
- [ ] K Nearest Neighbors
- [ ] Regularized Logistic Regression

Question 50

0.2 / 0.2 pts

Asking a thousand people hundreds of questions about their personalities, you can use which technique to find numbers which may approximate the "Big 5" personality characteristics.

Clustering

Classification

Correct!

Dimensionality reduction

'ou Answered

Regression

Quiz Score: **9.5** out of 10