Concept Quiz Over Week 1 Material

Due Oct 8 at 11:59pm **Points** 1 **Questions** 15
Available Oct 4 at 12am - Oct 8 at 11:59pm **Time Limit** None

Instructions

This quiz covers things we discussed in lecture during week one (and a bit of week 0).

Score for this survey: **1** out of 1
Submitted Oct 8 at 10:02am
This attempt took 75 minutes.

Question 1

In supervised learning, each data point in the training set has both an input representation and a known output. In unsupervised learning, no known outputs are provided.

ou Answered

- (selected) True
- False

True. Supervised learning assumes the data comes with correct output values annotated.

Question 2

You want to predict whether someone will like a particular music artist given a list of artists they already enjoy. You have a dataset of "liked artist lists", "like this new artist" pairs. This is an example of:

ou Answered

- Supervised Classification
- Unsupervised Clustering
- Supervised Regression
- Unsupervised Dimensionality Reduction

This is a supervised classification problem because we are trying to predict an outcome with a finite set of possible values (liked or not liked in this case). We have a dataset of "liked lists" and whether a new artist is liked as supervision.

Question 3

What is a hypothesis set?

ou Answered

- The set of possible functions for a machine learning algorithm
- A set of likely outputs for an input
- A set of hypothesis we have about what hyperparameters will work will for a problem

The set of possible functions for a machine learning algorithm. For instance, all possible weights for a linear regression model.

Question 4

What is meant by overfitting and underfitting?

Your Answer:

Over-fitting is when you take your data set and fit your parameters exactly to maximize accuracy on your data set. You are going to get a very high accuracy on your actual data set, but then, when you test your machine learning algorithm on other data sets, the accuracy is not going to be good because your function is hyper-specialized to the exact data set which you had trained on. An example of overfitting is a high-order polynomial function. This high order polynomial function will do a very good job of creating a complex function that fits every single point in the data set almost exactly, but the function is wildly varying, and when new points are tested with the function, the predictions will be way off. (high train accuracy, low test accuracy)

Under-fitting is when the model is not sufficiently complex enough to adequately capture the underlying complexity of the phenomenon which your machine learning algorithm is trying to capture. An example of this would be a function, which is linear, while the actual trend underlying the data set is squared. No matter how you shift or change the slope of the linear function, you will not be able to reasonably model the squared function. (low train and low test accuracy)

Overfitting is when performance on test data is significantly worse than performance on training data. Underfitting is when performance on both train and test is poor.

Question 5

Match the sources of error in function approximation with their definitions and remedies.

ɔu Answered **Modelling Error** Error from a mismatch

ɔu Answered **Estimation Error** Error from learning a r

ɔu Answered **Optimization Error** Error due to the difficu

ɔu Answered **Bayes Error** Irreducible error inheri

Modelling Error -> Error from a mismatch between our hypothesis set and the real function. Can be reduced with more expressive model classes.

Estimation Error -> Error from learning a model from finite dataset. Can be reduced with more data or with more data efficient algorithms

Optimization Error -> Error due to the difficulty of finding optimal models for a dataset during learning. Can be reduced with more computation dedicated to the search.

Bayes Error -> Irreducible error inherit to the problem. We can't really fix this.

Question 6

K-Nearest Neighbors is referred to as a _________ algorithm because no parameters are learned during training.

- Instance-based
- Exemplar

ou Answered

Non-Parametric

Non-parametric. As it does not learn any parameters.

Question 7

Parameters are parts of a model learned from data. Hyperparameters are settings for a model that are (generally) chosen by the machine learning practitioner.

ou Answered

True

False

True. Hyperparameters are set by the ML practitioner. Parameters are learned from data.

Question 8

Explain what a decision boundary is in classification.

Your Answer:

In classification, the area is broken up into different classes. Each point on the graph will represent a particular class. On the graph, there will be a specific location (for 2d, it will be a line) that changes from one class to another, this is called a decision boundary. Points on one side of the decision boundary will be classified as a different class to points on the other side of the decision boundary. This can be abstracted into higherorder spaces with no problem, for example, the decision boundary of a three-dimensional input space would be a surface, instead of a line.

A decision boundary is defined by the set of points in space where a classifier predicts equal confidence for all classes. For example, in logistic regression it is the set of points where w.Tx=0.

Question 9

List three hyperparameters for the k-Nearest Neighbor algorithm.

Your Answer:

k

distance metric (do you use Euclidean or Manhatten, or something else)

weighting function

k. Distance Function. Weighting Function.

Question 10

K-Fold Cross Validation can only be applied to k-Nearest Neighbors models.

ou Answered

- ( ) True
- (•) False

False. K-Fold Cross Validation can be applied to any learning algorithm.

Question 11

Each value in a probability density function must be less than 1.

ɔu Answered

- (selected) True
- False

Question 12

What does the IID (independently and identically distributed) assumption assume about our data?

- No answer text provided.
- That each data point comes from a different distributions.
- That each data point is identical.

ɔu Answered

- (selected) That data points do not affect other data points and all data points are generated from the same probabilistic mechanism.

That data points do not affect other data points and all data points are generated from the same probabilistic mechanism.

The "independent" part implies that data points do not affect each other and the "identically distributed" bit implies they are generated from the same distribution.

Question 13

What is Maximum Likelihood Estimation (MLE)?

Your Answer:

Maximum Likelihood Estimation starts by assuming that the model follows a probabilistic model. We therefore try to adjust the parameters of our models such that those parameters agree with our data set as much as possible. We say there is a particular probability of seeing our data set, and we try to adjust the parameter such that we maximize the probability of seeing our dataset.

Maximum Likelihood Estimation (MLE) is a way to fit parameters of a probabilistic model to data. In MLE, we assume some generative model of our data (i.e. a probabilistic model of how the data is produced) and then find parameters for that model that maximize the likelihood of our observed data. MLE is a general technique and we've now seen it applied to binary random variables (with a Bernoulli assumption), continue values (with a Normal assumption), and in linear regression (a conditional Normal assumption).

Question 14

When deriving MLE estimates in lecture, we frequently would write out the likelihood as:

$$L(\theta) \; = \; P(D \mid \theta) = \prod_{i=1}^{n} \; P(x_i \mid \theta)$$

What assumption allows us to write the probability of our dataset as a product of the probabilities of each data point?

- The Gaussian Assumption
- The Linearity of Expectation

ou Answered

- The IID Assumption

The IID assumption.

Independence between datapoints lets us write the joint probability of the dataset as the product of independent probabilities. If A and B are independent random variables, P(A,B) = P(A)P(B)

Question 15

Over all possible datasets generated from the true model, an estimator with high bias but low variance will _______________.

- Produce very different values for different datasets but will be correct on average.
- Consistently produce low-error estimates.

ou Answered

produce similar estimates for most datasets, but will not perfectly predict the true parameters.

produce similar estimates for most datasets, but will not perfectly predict the true parameters.

Bias refers to how wrong the provided estimate is on average. Variance refers to how much the predicted estimate varies across different datasets.

Survey Score: **1** out of 1