Score for this quiz: **83.4** out of 100

Submitted Oct 10 at 1:24am

This attempt took 47 minutes.

**Question 1**

**0 / 2 pts**

One of the powerful advantages of the Boosting algorithm is that it does not overfit.

**You Answered**

True

**Correct Answer**

False

**Question 2**

**0 / 2 pts**

Despite superficial similarities as graphs, Bayesian networks, and dependency trees are otherwise unrelated.

**You Answered**

True

**Correct Answer**

False

**Question 3**

**0 / 2 pts**

The output of a boosting algorithm that learns using “decision stumps” (i.e., a decision tree with only one node) can be converted to an equivalent ordinary decision tree in a straightforward way.

**You Answered**

True

**Correct Answer**

False

**Question 4**

**0 / 2 pts**

Your target concept is an element of your hypothesis space.

**Correct Answer**

True

**You Answered**

False

**Question 5**

**2 / 2 pts**

In general, when choosing a hypothesis space, a very large hypothesis space is preferable to a smaller one.

True

**Correct!**

False

**Question 6**

**2 / 3 pts**

What are potential issues with very deep decision trees?

**Correct!**

Overfitting to training data

Being insensitive to feature scaling

Underfitting due to simplicity

**Correct Answer**

Long computation times during predictions

**Correct!**

Reduced interpretability

Always providing the best accuracy

**Question 7**

**3 / 3 pts**

In the context of decision trees, which statements are true about leaf nodes?

**Correct!**

They represent an outcome or class label.

**Correct!**

They signify no further splitting is necessary for the instances reaching that node.

They represent root nodes.

They determine which attribute to split on next.

They are always binary.

They are always at the same depth.

**Question 8**

**1 / 3 pts**

Why might pruning be applied to a decision tree?

**Correct!**

To simplify the model and improve interpretability

To always achieve the best accuracy

**Correct!**

To reduce overfitting

To increase tree depth

**You Answered**

To ensure the tree is balanced

**Correct Answer**

To remove branches that provide little to no predictive power

**Question 9**

**3 / 3 pts**

Which algorithms are primarily used for classification tasks?

**Correct!**

Decision Trees

Polynomial Regression

**Correct!**

Support Vector Machines (SVM)

LASSO Regression

Ridge Regression

Linear Regression

**Question 10**

**2.4 / 3 pts**

When evaluating a classification model, which metrics can provide insights into its performance?

**Correct!**

Accuracy

**Correct!**

Precision

**Correct Answer**

Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

**Correct!**

F1-Score

R-squared

**Correct!**

Recall

**Question 11**

**3 / 3 pts**

Which of the following are key assumptions made by linear regression models?

**Correct!**

Linearity between features and the target variable

The model must contain at least three predictors

The target variable is categorical

Classes are well-separated

Features are clustered

**Correct!**

Homoscedasticity of residuals (constant variance of residuals)

**Question 12**

**1.5 / 3 pts**

What techniques can be used to improve the generalization of neural networks?

Decision tree pruning

Feature scaling using SVM

One-hot encoding of features

**Correct!**

Early stopping

Decision tree boosting

**Correct Answer**

Dropout

**Question 13**

**1 / 3 pts**

In the context of neural networks, which of the following can help in preventing overfitting?

**Correct Answer**

Adding dropout layers

Use k-means clustering

**You Answered**

Increasing learning rate

One-hot encoding

**Correct!**

Regularization (e.g., L2 regularization)

**Correct!**

Data augmentation

**Question 14**

**3 / 3 pts**

When considering the number of parameters in a neural network, which of the following statements are true?

The number of parameters is always equal to the number of neurons.

**Correct!**

The number of parameters can influence overfitting, with more parameters often increasing the risk.

Parameters are solely associated with activation functions, not weights or biases.

The number of parameters is only influenced by the size of the input layer.

**Correct!**

Deep networks with many layers generally have more parameters.

Networks with more parameters always perform better on unseen data.

**Question 15**

**1.5 / 3 pts**

What considerations are essential when selecting the value of 'k' in k-Nearest Neighbors (KNNs)?

**Correct!**

A smaller 'k' can lead to a noisier model with higher variance

'k' should be directly proportional to the learning rate for optimal results

The value of 'k' determines the activation function used in the model

'k' should always be set to the square root of the number of features

The choice of 'k' should always be an even number to avoid ties

**Correct Answer**

A larger 'k' can provide smoother boundaries but might be computationally more expensive

**Question 16**

**3 / 3 pts**

How does the performance of instance-based learners like KNNs typically change as we add more training data?

The model starts to discard older data automatically.

The algorithm's training phase becomes much slower.

The algorithm becomes less sensitive to the choice of k.

**Correct!**

Query time (prediction time) generally increases.

**Correct!**

The model can better generalize to new data with increased training instances.

The distance metric becomes less relevant.

**Question 17**

**2 / 3 pts**

What are common distance metrics used in instance-based learning, such as k-Nearest Neighbors (KNN)?

Gini impurity

Cross-entropy loss

Mean squared error (MSE)

**Correct Answer**

Minkowski distance

**Correct!**

Euclidean distance

**Correct!**

Manhattan distance

**Question 18**

**3 / 3 pts**

For boosting algorithms like AdaBoost, which of the following are characteristic features?

They operate primarily on the principle of diversity through data subsetting

They involve random feature selection for each learner

**Correct!**

They combine weak learners sequentially to form a strong learner

They always use decision trees with a depth greater than 10

**Correct!**

They adjust the weights of misclassified instances to focus on them in subsequent models

They require normalization of data before training

**Question 19**

**3 / 3 pts**

Why might one use ensemble learning techniques?

**Correct!**

To combine multiple models' strengths and mitigate individual weaknesses

**Correct!**

To prevent overfitting by leveraging diversity

To handle missing values in the data

To provide a more interpretive model

To speed up training times for large datasets

**Correct!**

To reduce variance and improve generalization

**Question 20**

**3 / 3 pts**

Which of the following best describe the principle behind bagging?

It primarily focuses on reducing bias in the final model

It requires a sequential training of models

It adjusts the weights of instances after every iteration

It uses only a single type of learner for all the ensembles

**Correct!**

It aims to reduce variance by averaging multiple predictions

**Correct!**

It involves training multiple models on bootstrapped samples of the data

**Question 21**

**2 / 3 pts**

When understanding the computational complexity of SVMs, which of the following statements are accurate?

SVMs, by design, are always more computationally efficient than neural networks

Training complexity is always O(n^2) regardless of the kernel used

**Correct Answer**

The number of support vectors can influence prediction time as predictions involve computations with the support vectors

The number of features in the dataset does not affect the training time of SVMs

**Correct!**

Quadratic programming is a central component of training SVMs, which involves solving optimization problems

**Correct!**

SVMs can become computationally intensive as the size of the dataset grows, especially with certain kernel methods

**Question 22**

**1 / 3 pts**

For the Radial Basis Function (RBF) kernel in SVMs, which considerations are important?

Large values of gamma always lead to underfitting the data

Gamma doesn't play any role in the RBF kernel and can be ignored

**Correct!**

The choice of the gamma parameter influences the flexibility of the decision boundary

**Correct Answer**

A small gamma will produce a more flexible decision boundary, possibly leading to overfitting

**You Answered**

The RBF kernel is equivalent to a polynomial kernel of infinite degree

**Correct!**

RBF kernel requires the data to be normalized before training for optimal performance

**Question 23**

**3 / 3 pts**

Which of the following are commonly used kernel functions in SVMs?

L1 regularization function

**Correct!**

Radial Basis Function (RBF) kernel

**Correct!**

Sigmoid kernel

**Correct!**

Polynomial kernel

Tanh activation function

ReLU kernel

**Question 24**

**2 / 3 pts**

Which of the following are true regarding Occam's razor principle in the context of machine learning?

**Correct Answer**

It is related to the bias-variance trade-off where simpler models might have higher bias but lower variance

It implies that the VC dimension of a hypothesis space should always be maximized

It states that models with the most parameters always perform best on unseen data

**Correct!**

It suggests a preference for simpler hypotheses over more complex ones

It discourages the use of kernel methods in SVMs

**Correct!**

It is based on the notion that simplicity often leads to better generalization

**Question 25**

**1 / 3 pts**

Regarding the bias-variance trade-off in the context of computational learning theory, which of the following are true?

Bias and variance are independent, and changing one does not affect the other

**Correct Answer**

Bias refers to the error introduced by approximating a real-world problem by a too-simple model

A model with high bias always has a low VC dimension

**Correct Answer**

Overfitting can be a result of too low bias and too high variance

High variance is always desirable as it ensures the model adapts well to the training data

**Correct!**

Variance refers to the error introduced by a model's sensitivity to small fluctuations in the training set

**Question 26**

**3 / 3 pts**

Which of the following are essential components of the PAC (Probably Approximately Correct) learning framework?

A fixed set of features to represent all possible inputs

A specific learning algorithm, such as a neural network or SVM

**Correct!**

An error measure representing the probability that a hypothesis will misclassify a randomly drawn instance

**Correct!**

A confidence parameter representing the probability that a hypothesis will perform worse than the error measure

**Correct!**

A hypothesis space from which hypotheses are drawn

**Correct!**

A sample complexity determining the number of examples required to achieve a certain error and confidence level

**Question 27**

**1 / 3 pts**

When considering linear classifiers in a 2D plane, which statements about their Vapnik–Chervonenkis (VC) dimension are true?

Linear classifiers can shatter any number of points given enough parameters

The VC dimension of linear classifiers decreases as the data dimensionality increases

**Correct Answer**

The VC dimension of a perceptron in 2D is 3

Linear classifiers can shatter any configuration of four distinct points in 2D

**Correct!**

If a hypothesis class has a VC dimension of d, there exists some set of d points that it can shatter

**Correct Answer**

The VC dimension provides a bound on the number of points that can be separated with a straight line for every possible labeling

**Question 28**

**3 / 3 pts**

In the context of polynomial classifiers in a 2D plane, how does the degree of the polynomial relate to the Vapnik–Chervonenkis (VC) dimension?

A 2nd-degree polynomial always has a VC dimension of 2.

**Correct!**

Higher-degree polynomials can capture more complex boundaries, potentially increasing the VC dimension.

The VC dimension remains constant regardless of polynomial degree.

Only odd-degree polynomials can have a VC dimension greater than 1.

**Correct!**

As the degree of the polynomial increases, the VC dimension generally increases, allowing the classifier to shatter more point configurations.

The degree and VC dimension are inversely related; as the degree goes up, the VC dimension goes down.

**Question 29**

**3 / 3 pts**

How does the Vapnik–Chervonenkis (VC) dimension of a hypothesis class impact its PAC learnability?

**Correct!**

The VC dimension provides a theoretical framework for understanding how complex a model can be while still being learnable from finite data.

PAC learnability is determined solely by the type of learning algorithm and not the VC dimension.

**Correct!**

The larger the VC dimension, the more training samples might be needed to ensure PAC learnability under the same confidence and accuracy constraints.

The VC dimension has no bearing on PAC learnability; all hypothesis classes are equally PAC learnable.

Only hypothesis classes with a VC dimension of 1 are PAC learnable.

**Correct!**

A finite VC dimension is a necessary condition for a hypothesis class to be PAC learnable.

**Question 30**

**3 / 3 pts**

What are key principles underlying Bayesian learning?

**Correct!**

It combines prior beliefs (prior probability) with evidence (data) to form a more refined belief (posterior probability).

**Correct!**

It applies the principle of probability to infer the best hypothesis given the data.

It assumes all hypotheses are equally probable a priori.

**Correct!**

It uses Bayes' theorem to update the probability estimate for a hypothesis as more evidence becomes available.

It always produces deterministic results for any given dataset.

It is exclusively used for regression problems.

**Question 31**

**3 / 3 pts**

In the context of Bayesian learning, which statements about the likelihood are correct?

It is the same as the prior probability for a hypothesis.

**Correct!**

The likelihood represents the probability of observing the data given a specific hypothesis.

It's always uniform across all hypotheses.

**Correct!**

Bayes' theorem uses the likelihood to weigh the evidence provided by the data.

A higher likelihood always indicates a more probable hypothesis.

**Correct!**

It quantifies how well a hypothesis explains the observed data.

**Question 32**

**3 / 3 pts**

When applying Bayesian learning, what role does the prior probability play?

The prior is determined by the current dataset, not any previous knowledge.

**Correct!**

It encodes any previous beliefs or knowledge about the hypotheses before observing the data.

In Bayesian learning, the prior is always biased towards the most complex hypothesis.

**Correct!**

The posterior probability is computed by updating the prior based on the data's likelihood.

**Correct!**

Priors can be uninformative (flat) when there's no prior knowledge or belief.

**Correct!**

Incorporating a prior allows the model to integrate domain knowledge into the learning process.

**Question 33**

**2 / 3 pts**

In Bayesian models, what can lead to the "overfitting" phenomenon?

**Correct!**

Neglecting to account for model complexity, leading to an overly tailored fit to the observed data.

Using a uniform prior for every parameter.

Always choosing the hypothesis with the maximum prior probability.

**Correct Answer**

Strong priors that don't align well with the actual data characteristics.

The models always produce results that are independent of the training data.

**Correct!**

Complex models with many parameters but insufficient data to robustly inform those parameters.

**Question 34**

**3 / 3 pts**

What advantages does Bayesian inference offer over traditional (frequentist) statistical methods?

**Correct!**

It results in a full probability distribution over parameters, capturing uncertainty.

It always results in simpler models with fewer parameters.

Bayesian methods are always faster and more computationally efficient.

**Correct!**

It offers a natural way to handle missing data or hierarchical structures.

It never requires assumptions about underlying data distributions.

**Correct!**

It can incorporate prior knowledge or beliefs into the analysis.

**Question 35**

**3 / 3 pts**

The Bayesian approach to machine learning provides a principled method for handling uncertainty. How is this achieved?

**Correct!**

By updating beliefs (or probabilities) in light of new data using Bayes' theorem.

By considering only the likelihood and disregarding prior information.

**Correct!**

By representing beliefs about parameters or hypotheses using probability distributions.

**Correct!**

Through the posterior distribution, which combines prior beliefs and the evidence from the data.

By avoiding any model that has a non-zero prior.

By always choosing the most probable hypothesis without considering alternatives.

Quiz Score: **83.4** out of 100

This quiz score has been manually adjusted by +10.0 points