

QUESTION 1

1 points

✓ Saved

..... is the process of converting an actual range of values that a numerical feature can take, into a standard range of values, typically in the interval $[-1, 1]$ or $[0, 1]$.

- ☐ Standardization
- ☐ Binning
- ☐ One-Hot Encoding
- ☒ Normalization

QUESTION 2

1 points  Saved

A typical machine learning workflow follows this order:

- ☐ Data Collection → Data Preparation → Model Selection & Training → Model Tuning → Model Evaluation → Model Deployment
- ☒ Data Collection → Data Preparation → Model Selection & Training → Model Evaluation → Model Tuning → Model Deployment
- ☐ Data Collection → Data Preparation → Model Selection & Training → Model Deployment → Model Evaluation → Model Tuning
- ☐ Data Collection → Data Preparation → Model Selection & Training → Model Deployment → Model Tuning → Model Evaluation

QUESTION 3

1 points  Saved

The principal hyperparameters to select and tune in k-Nearest Neighbors are:

- ☐ The value for k and the choice of the kernel function
- ☒ The value for k and the choice of the distance metric
- ☐ The choice of the activation function and the choice of the kernel function
- ☐ The value for k and the choice of the activation function

QUESTION 4

1 points

✓ Saved

Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?

- ☐ Decision Tree
- ☐ Kernel Regression
- ☐ Support Vector Machine
- ☒ Random Forest

QUESTION 5


1 points

Save Answer

The objective of the algorithm is to find a hyperplane in an N-dimensional space (N is the number of features) that distinctly classifies the data points.

- ☒ Support Vector Machine
- ☐ Kernel Regression
- ☐ k-Nearest Neighbors
- ☐ Logistic Regression

QUESTION 6

1 points  Saved

To assess a regression model, we can simply use a metric like:

- ☐ Confusion matrix
- ☐ Accuracy
- ☐ F1 score
- ☒ Mean Squared Error

QUESTION 7

1 points

✓ Saved

..... is an example of a frequently used optimization algorithm for finding the minimum of a function, and it is used in cases where the optimization criterion is differentiable.

- ☐ RBF kernel
- ☒ Gradient Descent
- ☐ Euclidean distance
- ☐ ReLU

QUESTION 8

1 points

Save Answer

Which one of these algorithms is an instance-based learning algorithms?

- ☐ Neural Networks
- ☒ Kernel Regression
- ☐ Linear Regression
- ☐ Logistic Regression

QUESTION 9

1 points

Save Answer

Some machine learning algorithms are less sensitive to the problem of an imbalanced dataset such as:

- ☐ Decision trees
- ☐ Random forest
- ☐ Gradient boosting
- ☒ All of the above

QUESTION 10

1 points

✔ Saved

An activation function in a neural network defines how the weighted sum of the input is transformed into an output from a node or nodes in a layer of the network. Which one of these is not an example of activation functions?

- ☐ TanH
- ☒ GMM
- ☐ ReLU
- ☐ Sigmoid

QUESTION 11


1 points

Save Answer

..... is a hold-out set used to choose the learning algorithm and to find the best values of hyperparameters:

- ☐ Test set
- ☒ Validation set
- ☐ Original data set
- ☐ Training set

QUESTION 12

1 points  Saved

The principal hyperparameters to tune in gradient boosting are:

- ☐ The number of trees, the learning rate, and the selection of an activation function
- ☒ The number of trees, the learning rate, and the depth of trees
- ☐ The learning rate, and the depth of trees
- ☐ The number of trees and the depth of trees

QUESTION 13


1 points

Save Answer

The idea of is that we start the learning with relatively few labeled examples, and a large number of unlabeled ones, and then add labels only to those examples that contribute the most to the model quality.

- ☐ Dimensionality reduction
- ☒ Active Learning
- ☐ Clustering
- ☐ Reinforcement learning

QUESTION 14

1 points  Saved

Which of the following is a disadvantage of decision trees?

- ☐ Decision trees output are hard to interpret and explain
- ☐ Decision trees are robust to outliers
- ☒ Decision trees are prone to be overfit
- ☐ None of the above

QUESTION 15

1 points

✓ Saved

Which one of these is often used in text processing and speech processing because sentences and texts are naturally sequences of either words/punctuation marks or sequences of characters?

- ☐ Convolutional Neural Network (CNN)
- ☒ Recurrent Neural Network (RNN)
- ☐ Feed-Forward Neural Network
- ☐ None of the above

QUESTION 16

1 points

✓ Saved

Which of these practices can be used to handle missing or corrupted data in a dataset?

- ☐ Drop missing rows or columns (features)
- ☐ Replace missing values with mean/median
- ☐ Assign a unique category/class to missing values
- ☒ All of the above

QUESTION 17


1 points



The key difference between L1 and L2 regularization is that:

- ☐ L2 shrinks the less important feature's coefficient to zero thus, removing some feature altogether
- ☐ There are no differences between them, they are identical
- ☒ L1 shrinks the less important feature's coefficient to zero thus, removing some feature altogether
- ☐ Both L1 & L2 shrink the less important feature's coefficient to zero but L2 can produce a negative feature's coefficient

QUESTION 18

1 points  Saved

There are several techniques when handling imbalanced datasets. A popular technique is:

- ☒ Synthetic Minority Oversampling Technique (SMOTE)
- ☐ Uniform Manifold Approximation and Projection (UMAP)
- ☐ Gaussian mixture model (GMM)
- ☐ Density-based spatial clustering of applications with noise (DBSCAN)

QUESTION 19

1 points

Save Answer

Several reasons can lead to the problem of overfitting, an important one is:

- ☒ The dataset has too many features but with a small number of training examples
- ☐ The model is too simple for the data
- ☐ The engineered features are not informative enough
- ☐ None of the above

QUESTION 20

1 points

Save Answer

The purpose of the function `predict()` in `sklearn` is to:

- ☐ evaluate the prediction results using cross validation on the data
- ☒ perform predictions on the testing instances, based on the learned parameters
- ☐ evaluate the prediction results using a random split of the data
- ☐ fit the model to the input training instances