**Self-Study Knowledge Check 1.1: Introduction to Machine Learning**

**Question 1**

**1 / 1 pts**

Machine learning is the study of all of the different ways in which models can be built from data.

**Correct!**



True

*You are correct! “True” is correct because machine learning encompasses all the ways models can be built to model systems.*



False

**Question 2**

**1 / 1 pts**

The more complex the task or problem becomes, (blank). *(Check all that apply.)*

**Correct!**



...the more computational power is needed to process that data.

*You are correct! All of these* *are correct because all the scenarios occur if the task or problem is complex.*

**Correct!**



...the more data is needed to train the model.

*You are correct! All of these* *are correct because all the scenarios occur if the task or problem is complex.*

**Correct!**



...the more complicated the model needs to be to represent it.

*You are correct! All of these* *are correct because all the scenarios occur if the task or problem is complex.*

**Question 3**

**1 / 1 pts**

In Python, a list of numbers can be represented by which of the following? *(Check all that apply.)*



Dataframe

**Correct!**



Array

*You are correct! In Python, a list of numbers can be represented with a vector, array, or tuple.*

**Correct!**



Tuple

*You are correct! In Python, a list of numbers can be represented with a vector, array, or tuple.*



String

**Correct!**



Vector

*You are correct! In Python, a list of numbers can be represented with a vector, array, or tuple.*

**Question 4**

**1 / 1 pts**

Your input is a bolded letter X. Because it is bolded, you know that it represents a(blank)of numbers?

**Correct!**



List

*You are correct! The answer “List” is correct because, for inputs, a bolded letter is used for a list of numbers.*



Char



Group



String

**Question 5**

**1 / 1 pts**

The predicted output of a function f(x) is represented by?



x



X̂

**Correct!**



ŷ

*You are correct! The symbol “ŷ” is correct because it is the symbol for predicted output.*



Y

**Question 6**

**1 / 1 pts**

The function which takes the actual result y and our predicted result ŷ and scores it as the square of the difference is known as the loss function L.

**Correct!**



True

*You are correct! The loss function (L) is calculated as the square of the difference between the actual result and the predicted result.*



False

# Self-Study Knowledge Check 1.2: Distribution and Random Variables

**Question 1**

**1 / 1 pts**

What do you call a system with the same input which returns a different output each time it is sampled?



Deterministic system

**Correct!**



Non-deterministic system

*You are correct! The answer “Non-deterministic system*” *is correct because such a system exhibits a different behavior every time it is sampled.*



Problematic system



Probabilistic system

**Question 2**

**1 / 1 pts**

A distribution is also known as a probability density function.

**Correct!**



True

*You are correct! The terms ‘distribution’ and ‘probability density function’ refer to the same thing.*



False

**Question 3**

**1 / 1 pts**

When a coin is tossed, there are two possible outcomes: heads or tails. What is this type of output known as?

**Correct!**



Discrete distribution

*That is correct! The answer “discrete distribution” is correct because in this system, the data can be only certain values such as integers, (0,1,2, and 3, for example).*



Continuous distribution



Wide-spread distribution



Predictable distribution

**Question 4**

**1 / 1 pts**

If the output of a function can be any real number, the distribution is known as a (blank).

**Correct!**



Continuous distribution

*You are correct! The answer “Continuous distribution“ is correct because in continuous distribution, the random variable X can have any value since there are infinite values X can take.*



Predictable distribution



Wide-spread distribution



Discrete distribution

**Question 5**

**1 / 1 pts**

The generic output of the system is represented with a capital letter Y and the samples of the output as small letter y.

**Correct!**



True

*You are correct! The answer “True” is correct because the capital letter Y is used to denote the generic output of a system, and the samples of the output are denoted with a small letter y.*



False

**Question 6**

**1 / 1 pts**

What is represented by Y~Fy?



Y is sort of like Fy.



Y is distributed according to a distribution y of F.

**Correct!**



Y is distributed according to a distribution Fy.

*You are correct! The answer “Y is distributed according to a distribution Fy” is correct because Y is sampled using the function Fy.*



Y is a function of F.

**Question 7**

**1 / 1 pts**

What is the size of the output of the formula [y]*12?*



3



13

**Correct!**



12

*You are correct! The answer “12” is correct because the subscript number in the question denotes how many unique outputs to return.*



11

# Self-Study Knowledge Check 1.4: Expected Value and Variance

**Question 1**

**1 / 1 pts**

In a continuous distribution, the area under the curve between any two vertical lines can never exceed what number?



Two



Zero

**Correct!**



One

*You are correct! The answer “One” is correct because the sum of the entire area under the curve will always equal one. Any area between two vertical lines running through the distribution will always be less than one.*

**Question 2**

**0.5 / 1 pts**

How can you represent the expected value of a random variable Y? *(Check all that apply.)*

**Correct Answer**



μy



E[X]



Xvar

**Correct!**



E[Y]

*You are correct! The answers “E[Y]” and “*μy*” are correct because these are the correct representations for the expected value of Y.*

**Question 3**

**1 / 1 pts**

The mean and median of a distribution can be the same.

**Correct!**



True

*You are correct! The answer “True” is correct because although the mean and median measure different things, they can be the same value.*



False

**Question 4**

**1 / 1 pts**

The law of large numbers says that the average will converge to the expectation of Y in the limit as n becomes large.

**Correct!**



True

*You are correct! The answer “True” is correct because as n grows, there are more data points to base the average (mean) calculations on, and therefore, it becomes a better representation of the expected value.*



False

**Question 5**

**1 / 1 pts**

If a distribution has a large variance its graph will be wider and shorter.

**Correct!**



True

*You are correct! The answer “True” is correct because variance is a measure of how widely distributed the values are. The larger the variance grows, the wider the graph will be.*



False

**Question 6**

**1 / 1 pts**

What is the formula for the standard deviation?



The sum of all samples divided by the number of samples

**Correct!**



The square root of the variance

*You are correct! The answer “The square root of the variance” is correct because standard deviation is the square root of the variance.*



The same as the variance



The cube root of the variance

**Question 7**

**1 / 1 pts**

What is the symbol to represent the standard deviation of a distribution?



Var[y]



𝛔y2



E[y]

**Correct!**



𝛔y

*You are correct! The answer “*𝛔y” *is correct because this is the correct symbol for the standard deviation.*

# Self-Study Knowledge Check 1.5: Introduction to pandas

**Question 1**

**1 / 1 pts**

Python numpy arrays are homogeneous.

**Correct!**



True

*You are correct! The answer “True” is correct because all numpy arrays are homogeneous.*



False

**Question 2**

**1 / 1 pts**

Select the homogenous array.



[‘b’,’x’,0,’y’]



[0,’a’,1,2,3]

**Correct!**



[0,1,2,3,4]

*That is correct! The answer “[0,1,2,3,4]” is correct because all the elements in the array are of same type, that is, integers.*



[0,0,0,0,’b’]

**Question 3**

**1 / 1 pts**

Each column in a dataframe is homogenous.

**Correct!**



True

*You are correct! The answer “True” is correct because each column in a dataframe must contain only one type of data.*



False

**Question 4**

**1 / 1 pts**

What is the standard alias used when importing a pandas library?



df



pandas



pand

**Correct!**



pd

*You are correct! The answer “pd” is correct because this is the standard alias.*

**Question 5**

**1 / 1 pts**

What code would you use to create a dataframe from a variable type dictionary stored as data?



DF=pd.dataframe.data

**Correct!**



DF=pd.dataframe(data)

*You are correct! The answer “*DF=pd.dataframe(data)*” is correct because the data definition is inside the parentheses of the dataframe function.*



DF=pd.dataframe”data”



DF=pd.dataframe’data’

**Question 6**

**1 / 1 pts**

When creating a dataframe, an index is created by default (if not provided).

**Correct!**



True

*You are correct! The answer “True” is correct because pandas will create an index if there is not one specified when the dataframe is created.*



False

**Question 7**

**1 / 1 pts**

How can an index be set in creating a new dataframe? *(Check all that apply.)*



Import an index from the Internet.

**Correct!**



Pass an index to the constructor with the index input argument.

*You are correct! “Pass an index to the constructor with the index input argument”, “Use*set\_index()*to assign a column to be the index”, and “Keep the default index” are all correct ways to set an index when creating a new dataframe.*

**Correct!**



Use set\_index() to assign a column to be the index.

*You are correct! “Pass an index to the constructor with the index input argument”, “Use*set\_index()*to assign a column to be the index”, and “Keep the default index” are all correct ways to set an index when creating a new dataframe.*

**Correct!**



Keep the default index.

*You are correct! “Pass an index to the constructor with the index input argument”, “Use*set\_index()*to assign a column to be the index”, and “Keep the default index” are all correct ways to set an index when creating a new dataframe.*

**Question 8**

**1 / 1 pts**

What is the syntax to load a .csv file into a dataframe?



df=pd.read\_csv’filename’



df=read\_csv(filename)

**Correct!**



df=pd.read\_csv(filename)

*You are correct! The answer “*df=pd.read\_csv(filename)*” is correct because the statement calls pandas (pd) as well as the function (read\_csv()) and passes the filename inside the function parentheses.*



df=pd.readcsv(filename)

**Question 9**

**1 / 1 pts**

Which function is used to get the datatypes of the columns in the dataframe?

**Correct!**



info()

*You are correct! The answer “*info()*” is correct because info() is used to get the datatype of each column in the dataframes.*



tail()



head()



describe()

**Question 10**

**1 / 1 pts**

Which of the following functions would show you count, mean, std, min, max?



info()

**Correct!**



describe()

*You are correct! The answer “*describe()*” is correct because*describe()*is used to get the statistical data analysis of the dataframe columns, that is, the mean, median, and std.*



tail()



head()

**Question 11**

**1 / 1 pts**

What will df.head(10) return?



The center 10 rows of the dataframe df



The last 10 rows of the dataframe df

**Correct!**



The top 10 rows of the dataframe df

*You are correct! The answer “The top 10 rows of the dataframe df” is correct because the function*head()*is used to print the starting rows of the data, and when given a parameter ‘10’, it will return the first (or top)10 rows.*



Ten random rows from the dataframe df

**Question 12**

**1 / 1 pts**

The function tail() is used to show the last rows of the dataframe.

**Correct!**



True

*That is correct! The answer “True” is correct because*tail()*is used to show the last rows of the dataframe.*



False

# Self-Study Knowledge Check 1.6: Selecting Data in pandas

**Question 1**

**1 / 1 pts**

Which of the following is a way to select data from a dataframe? *(Check all that apply.)*

**Correct!**



table.iloc[]

*You are correct! The answers “*table[]*”, “*table.loc[]*”, and “*table.iloc[]”*are all correct because all of them are statements to select data from a dataframe.*

**Correct!**



table[]

*You are correct! The answers “*table[]*”, “*table.loc[]*”, and “*table.iloc[]”*are all correct because all of them are statements to select data from a dataframe.*



table.data()

**Correct!**



table.loc[]

*You are correct! The answers “*table[]*”, “*table.loc[]*”, and “*table.iloc[]”*are all correct because all of them are statements to select data from a dataframe.*

**Question 2**

**1 / 1 pts**

Suppose you have a dataframe:

{’X1’:[0,1,2]

 ‘Y1’:[’a’,’b’,’c’]

 ‘Y’ :[0.1,0.2,0.3]

}

To select column X1 and Y1 into a new dataframe, what is the correct syntax?



df= df[’X1’,’Y1’]



df= df[[X1,’Y1’]]

**Correct!**



df= df[[’X1’,’Y1’]]

*You are correct! The answer “*df= df[[’X1’,’Y1’]]*” is correct because this is the correct statement to select the columns X1 And Y1.*



df= df[[’X1’,Y1]]

**Question 3**

**1 / 1 pts**

If you want to select the first four rows of a dataframe, how can you do this using slice indexing?



DF=pd.head(4)



DF[4:]

**Correct!**



DF[:4]

*You are correct! The answer “*DF[:4]*” is correct because this code tells pandas to select rows from zero to 4 using slice indexing. If you leave the space before the colon blank, pandas assumes you want to start at the first row.*



DF[4]

**Question 4**

**1 / 1 pts**

What will the output of DF[DF[’meters’]>2] be?



Return only one column (meters) and only rows where the data in the column is greater than 2.

**Correct!**



Return all the rows and columns where data in the column ‘meters’ is greater than 2

*That is correct! The answer “Return all the rows and columns where data in the column ‘meters’ is greater than 2” is correct because the statement has a condition which states meters > 2.*



Return at least 2 columns.



Return rows where the data in the column ‘meters’ is equal to 2.

**Question 5**

**1 / 1 pts**

In which scenario is .loc[] not preferable to []?

**Correct!**



.loc[slice] when the index is non integers

*You are correct! The answer “*.loc[slice]*when the index is non-integers” is correct because it is only when slicing the index is non-integers that the function*.loc[]*is not applicable.*



.loc[slice] when the index is integers



.loc[slice] when the index is real numbers

**Question 6**

**1 / 1 pts**

What is the output of .loc[[5,6,9]]?

**Correct!**



The output would be three rows of data: rows with the index of 5, 6, and 9.

*You are correct! The answer “The output would be three rows of data: rows with the index of 5, 6, and 9” is correct because the statement has a function .loc[] with defined indexes which are to be returned.*



The output would be rows greater than 5, 6, and 9.



The output would be columns with the names 5, 6, and 9.

**Question 7**

**1 / 1 pts**

The statement .loc[0:3,[’column1’,’column4’]] will return data from indexes 0 to 3, specific to column1 and column4.

**Correct!**



True

*You are correct! The answer "True" is correct because the statement will return data from rows with indexes from 0 to 3, but specifically for column1 and column4 only, not the columns in between (column2 and column3).*



False

**Question 8**

**1 / 1 pts**

For .iloc[], the row selector and the column selectors can be integers, list of integers, or slices.

**Correct!**



True

*You are correct! The answer “True” is correct because the row selector can be any of these three.*



False

**Question 9**

**1 / 1 pts**

What is the difference between .loc[] and .iloc[]? *(Check all that apply.)*

**Correct!**



.iloc[] can have slicing in the column selector

**Correct!**



.iloc[] can have integers in the column selector

**Correct!**



.iloc[] can have a list of integers in the column selectors

*You are correct! The answers “*.iloc[]*can have slicing in the column selector”’, “*.iloc[]*can have integers in the column selector”, and “*.iloc[]*can have a list of integers in the column selectors” are all correct.*

# Self-Study Knowledge Check 1.7: pandas Operations and Plots

**Question 1**

**1 / 1 pts**

With the dataframe df= pd.dataframe({’A’:[0.1,0.2,0.3], ‘B’:[10,20,30]}), what does the statement df[’B’].sum() return?

**Correct!**



60

*You are correct! The answer “60” is correct because the statement returns the sum of the column ‘B’ which adds up to 60.*



30



20



10

**Question 2**

**1 / 1 pts**

With the dataframe df= pd.dataframe({’A’:[0.1,0.2,0.3], ‘B’:[10,20,30]}), to add column A and B, what is the correct syntax?

**Correct!**



df[’A’].add(df[’B’])

*You are correct! The answer “*df[’A’].add(df[’B’])*” is correct because to add two columns, you specify the first column, and then use .add() to specify the second column to add.*



df[A].add(df[B])



df[’A’].add.df[’B’]



df[’A’]add(df[’B’])

**Question 3**

**1 / 1 pts**

When using .plot(), the parameter ‘kind’ tells what type of plot is to be drawn.

**Correct!**



True

*You are correct! The answer “True” is correct because ‘kind’ is used to define what type of plot to create.*



False

**Question 4**

**1 / 1 pts**

With the dataframe df= pd.dataframe({’A’:[0.1,0.2,0.3], ‘B’:[10,20,30], ‘C’:[1,2,3]}), how would you create a histogram that includes columns B and C?



df.plot(kind = ‘scatter’,y=[’B’,’C’])

**Correct!**



df.plot(kind = ‘hist’,y=[’B’,’C’])

*That is correct! The answer “*df.plot(kind = ‘hist’,y=[’B’,’C’])*” is correct because this is the right method to build a histogram between column B and C.*



df.plot(kind = ‘hist’,y=[’B’,’A’])

**Question 5**

**1 / 1 pts**

For time series data, the best visualization is done using a line chart.

**Correct!**



True

*You are correct! The answer “True” is correct because time series data (typically) tracks data over time. To show the change over time, a line plot is the most appropriate method.*



False

**Question 6**

**1 / 1 pts**

With the dataframe df, to plot a line chart with columns A and B, what is the correct code?

**Correct!**



df.plot(kind=’line’,x=’index’,y=[’A’,’B’])

*You are correct! The answer “*df.plot(kind=’line’,x=’index’,y=[’A’,’B’])*” is correct because it sets the x-axis as the index and plots the lines of column A and B on the y-axis with the correct syntax.*



df.plot(kind=’line’,y=’index’,y=[’A’,’B’])



df.plot(kind=’line’,x=’index’,x=[’A’,’B’])

**Question 7**

**1 / 1 pts**

In the plot() function, what does the parameter figsize=(1,2) specify?



The rows of data to plot



The height of the plot

**Correct!**



The width and height of the plot

*You are correct! The answer “The width and height of the plot” is correct because the*figsize*parameter has two parts to set: the width and height.*



The width of the plot

**Question 8**

**1 / 1 pts**

With the dataframe df= pd.dataframe({’A’:[0.1,0.2,0.3], ‘B’:[10,20,30], ‘C’:[1,2,3]}), how can you create a bar chart?



df.plot(kind=’bar’)



df.sum().(kind=’bar’)

**Correct!**



df.sum().plot(kind=’bar’)

*You are correct! The answer “*df.sum().plot(kind=’bar’)*” is correct because a bar chart is plotted among the sum of the columns as bar heights.*



df.sum().plot(kind=bar)