**Self-Study Knowledge Check 8.1: Parabolic Model Fitting and Nonlinear Features**

**Question 1**

**1 / 1 pts**

If there are three features, the linear regression model predicts an output  �^= (blank).



*θ1* + *θ2* + *θ3*

**Correct!**



*θ1 Φ1* + *θ2 Φ2* + *θ3 Φ3*

*You are correct! The answer “θ1 Φ1* + *θ2 Φ2* + *θ3 Φ3” is correct because this is the formula for a three-feature linear regression model.*



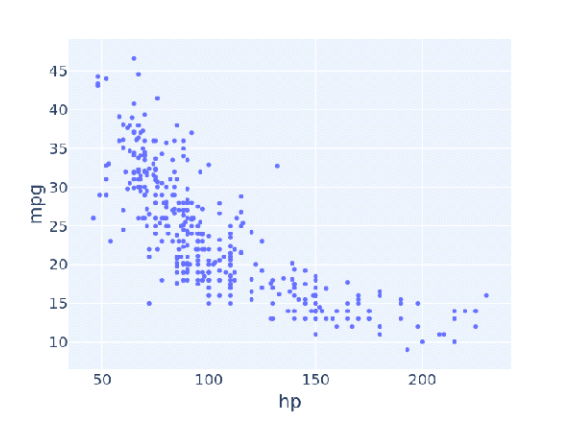
*Φ1* + *Φ2* + *Φ3*



*θ1 Φ2* + *θ2 Φ3* + *θ3 Φ1*

**Question 2**

**1 / 1 pts**



The graph shows that as horsepower or ‘hp’ increases the miles per gallon or ‘mpg’ increases.



True

**Correct!**



False

*You are correct! The answer “*False*” is correct because the graph shows that with the increase in horsepower or ‘hp’, the miles per gallon or ‘mph’ decreases.*

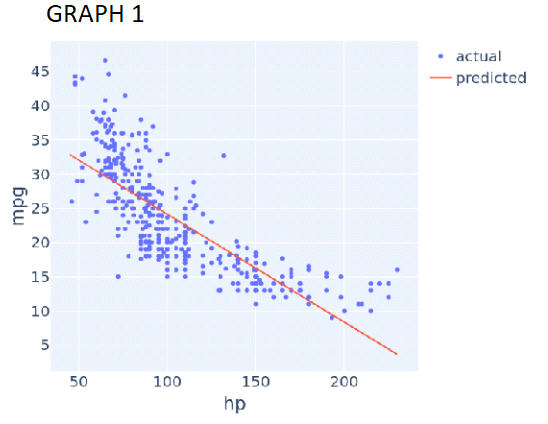
**Question 3**

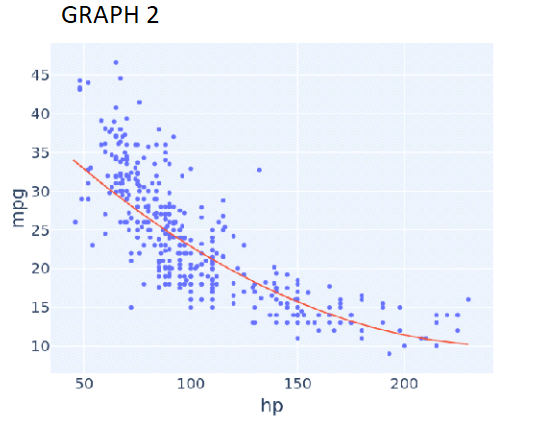
**1 / 1 pts**

Which of the following graphs is an example of a parabola approach?



GRAPH 1

**Correct!**

****



GRAPH 2

*You are correct! The answer “*GRAPH 2*” is correct because the trend line in the graph is curved.*

**Question 4**

**1 / 1 pts**

A simple linear regression model can have squared terms.



True

**Correct!**



False

*You are correct! The answer “*False*” is correct because a simple linear regression model cannot have squared terms.*

**Question 5**

**1 / 1 pts**

How do you include squared terms in a linear regression model?



Taking the square of the feature in the linear regression model equation



Using the SquaredRegression library from sklearn

**Correct!**



Adding a new column with squared values of the feature

*You are correct! The answer “*Adding a new column with squared values of the feature*” is correct because to include squared terms in a linear regression model, the solution is to add a new squared column of that feature.*

**Question 6**

**1 / 1 pts**

The library in Python “from sklearn.linear\_model import Linear regression” can be used for squared regression model building.

**Correct!**



True

*You are correct! The answer “*True*” is correct because it can be used for a squared regression model introducing squared terms of features.*



False

**Self-Study Knowledge Check 8.2: Scikit-Learn Transformers**

**Question 1**

**1 / 1 pts**

A transformer takes a set of existing features as input and outputs new features.

**Correct!**



True

*You are correct! The answer “True” is correct because the transformer takes a set of existing features as input and outputs new features.*



False

**Question 2**

**1 / 1 pts**

In PolynomialFeatures(degree), ‘degree’ is used to declare the square size.



True

**Correct!**



False

*You are correct! The answer “*False*” is correct because the constructor ‘degree’ of the polynomial is used to control the number of features added.*

**Question 3**

**1 / 1 pts**

A PolynomialFeatures object with the ‘degree’ constructor equal to ‘5’ is written as (blank).

**Correct!**



PolynomialFeatures(degree=5)

*You are correct! The answer “*PolynomialFeatures(degree=5)*” is correct because this is the syntax in Python to create a “PolynomialFeatures” object with the “degree constructor equal to 5”.*



PolynomialFeatures(“degree”=5)



PolynomialFeature(degree=5)

**Question 4**

**1 / 1 pts**

“PolynomialFeatures().fit\_transform()” will output a (blank).



List



Character



Pandas dataframe

**Correct!**



Numpy array

*You are correct! The answer “*Numpy array*” is correct because the output of “*PolynomialFeatures().fit\_transform()*” is a “Numpy array”.*

**Question 5**

**1 / 1 pts**

What is the output of the function “.get\_feature\_names\_out()” for the “PolynomialFeatures” object, given the feature name as “hp” and the degree as ‘3’?

**Correct!**



[“hp”,”hp2”,”hp3”]

*You are correct! The answer “*[“hp”,”hp2”,”hp3”]*” is correct because this is the output of the function to get the feature names.*



[“hp”,”hp2”,”hp3”]



[“hp3”]



[“hp”,”hp2”]

**Self-Study Knowledge Check 8.3: Scikit-Learn Pipelines**

**Question 1**

**1 / 1 pts**

Consider the following linear regression pipeline model with a polynomial transform degree equal to 3:

**vehicle\_data\_with\_cubic\_features =   
             pd.DataFrame(poly\_transform.fit\_transform(vehicle\_data[["hp"]]),  
             columns = poly\_transform.get\_feature\_names\_out())  
cu\_model.fit(vehicle\_data\_with\_cubic\_features, vehicle\_data[["mpg"]])**

The Python statement for the prediction of a 100-horsepower vehicle would be  
cu\_model.predict([[100]]).



True

**Correct!**



False

*You are correct! The answer “*False*” is correct because the Python statement for the prediction of a 100-horsepower vehicle would be*  
cu\_model.predict([[100,10000,1000000]]).

**Question 2**

**0.5 / 1 pts**

Consider the following linear regression pipeline model with a polynomial transform degree equal to 3:

**vehicle\_data\_with\_cubic\_features =**  
**pd.DataFrame(poly\_transform.fit\_transform(vehicle\_data[["hp"]]),**  
**columns = poly\_transform.get\_feature\_names\_out())**  
**cu\_model.fit(vehicle\_data\_with\_cubic\_features, vehicle\_data[["mpg"]])**

What would the Python statement for the prediction of a 100-horsepower vehicle be?

**Correct!**



cu\_model.predict([[100,10000,1000000]])

*You are correct! The answer “*cu\_model.predict([[100,10000,1000000]])*" is correct because it is a correct Python statement for the prediction of a 100-horsepower vehicle.*

**Correct Answer**



cu\_model.predict(poly\_transform.fit\_transform[[100]])



cu\_model.predict([[100]])



cu\_model.predict([[100\*\*3]])

**Question 3**

**1 / 1 pts**

Scikit-learn’s pipeline class is a useful tool for encapsulating multiple different transformers alongside an estimator into one object.

**Correct!**



True

Yo*u are correct! The answer “*True*” is correct because pipeline class is a useful tool for encapsulating multiple different transformers alongside an estimator into one object.*



False

**Question 4**

**1 / 1 pts**

Consider the following linear regression pipeline model:

**from sklearn.pipeline import Pipeline**  
**pipelined\_model = Pipeline([**  
**('josh\_transform', PolynomialFeatures(degree = 3, include\_bias = false)),**  
**('josh\_regression', LinearRegression())**  
**])**  
**pipelines\_model.fit(vehicle\_data[["hp"]], vehicle\_data["mpg"])**

The Python statement for the prediction of a 100-horsepower vehicle would be pipelined\_model.predict([[100]]).

**Correct!**



True

*You are correct! The answer “*True*” is correct because this is the correct statement used for the prediction of a pipelined model.*



False

**Self-Study Knowledge Check 8.4: Order 0 Through 6 Models on Vehicle Data**

**Question 1**

**1 / 1 pts**

Consider the following pipelined regression model:

**from sklearn.pipeline import Pipeline**  
**pipelined\_model = Pipeline([**  
**('josh\_transform', PolynomialFeatures(degree =  , include\_bias = False)),**  
**('josh\_transform', LinearRegression())**  
**])**  
**pipelined\_model.fit(vehicle\_data[["hp"]], vehicle\_data["mpg"])**

For a model to be of order six, what should be the value (highlighted) of the constructor degree?



five



six

**Correct!**



6

*You are correct! The answer “*6*” is correct because the constructor should be set to value “6” for a model to be of order six.*



5

**Question 2**

**1 / 1 pts**

Increasing model complexity seems to decrease the error.

**Correct!**



True

*You are correct! The answer “*True*” is correct because increasing the model’s complexity seems to decrease the error.*



False

**Question 3**

**1 / 1 pts**

Increasing model complexity seems to decrease the variance.



True

**Correct!**



False

*You are correct! The answer “*False*” is correct because increasing the model’s complexity seems to increase the variance.*

**Question 4**

**1 / 1 pts**

Being overly sensitive to data is known as overfitting.

**Correct!**



True

*You are correct! The answer “*True*” is correct because being overly sensitive to data is known as overfitting.*



False

**Self-Study Knowledge Check 8.5: The Dangers of Overfitting**

**Question 1**

**1 / 1 pts**

If there are N data points, then a degree N-1 model will also have an MSE equal to 1.



True

**Correct!**



False

*You are correct! The answer “*False*” is correct because if there are N data points, then a degree N-1 model will also have an MSE equal to 0.*

**Question 2**

**1 / 1 pts**

If there are four data points, there can always be a four-parameter model, three coefficients, and one (blank) that fit perfectly.



Constant

**Correct!**



Intercept

*You are correct! The answer “*Intercept*” is correct because in a four-parameter model, there are always three coefficients and one intercept that fit perfectly.*



Zero value



Coefficient

**Question 3**

**1 / 1 pts**

Suppose that the number of numerical features available exceeds the number of data points. If a linear regression model is used, there will be zero training error.

**Correct!**



True

*You are correct! The answer “*True*” is correct because when the number of numerical features available exceeds the number of data points in a linear regression model, there will be zero training error.*



False

**Self-Study Knowledge Check 8.6: Overfitting and Validation**

**Question 1**

**1 / 1 pts**

The cross-validation technique is used to detect the degree of the model build.



True

**Correct!**



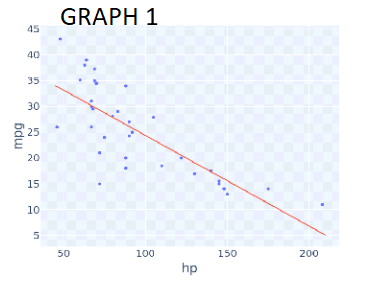
False

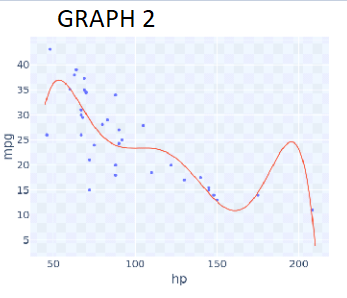
*You are correct! The answer “*False*” is correct because the cross-validation technique is used to detect the overfitting of the model build.*

**Question 2**

**1 / 1 pts**

Which of the following graphs seems to be overfitting?





**Correct!**



GRAPH 2

*You are correct! The answer “*GRAPH 2*” is correct because it can be seen that the curve of the graph is passing through a lot of data points and has a wiggly shape.*



GRAPH 1

**Question 3**

**1 / 1 pts**

In simple cross-validation the data is divided into which two sets?

**Correct!**



Development set

*You are correct! The answer “*Development set*” is correct because in simple cross-validation, it is one of the two datasets into which this data can be divided.*



Null set

**Correct!**



Training set

*You are correct! The answer “*Training set*” is correct because in simple cross-validation, it is one of the two datasets into which this data can be divided.*



Test set

**Question 4**

**1 / 1 pts**

The development set is used to compare models before they are trained.



True

**Correct!**



False

*You are correct! The answer “*False*” is correct because the development set is used to compare models after they are already trained.*

**Question 5**

**1 / 1 pts**

Consider the following code:

**from sklearn.utils import shuffle**  
**training\_set, dev\_set = np.split(shuffle(vehicle\_data\_sample\_35), [25])**

Given this code, the variable training\_set will have \_\_ rows if the total data size is 35.



10



15



35

**Correct!**



25

*You are correct! The answer “*25*” is correct because in the code, the initial division point defined is [25].*

**Question 6**

**1 / 1 pts**

What is the common shape of the validation error curve?

**Correct!**



Bowl shaped

*You are correct! The answer “*Bowl shaped*” is correct because the shape of the validation error curve is mostly bowl shaped.*



T shaped



L shaped



Straight lined

**Question 7**

**1 / 1 pts**

Hyperparameters are parameters whose values control the learning process.

**Correct!**



True

*You are correct! The answer “*True*” is correct because hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning.*



False

**Self-Study Knowledge Check 8.7: Test Sets**

**Question 1**

**1 / 1 pts**

The training set obtained from the split of the datasets is used for the selection of hyperparameters.



True

**Correct!**



False

*You are correct! The answer “*False*” is correct because training sets are used to train models of various complexity.*

**Question 2**

**1 / 1 pts**

In machine learning, the final estimate of outcome is ideally done on a third dataset. What is this third dataset commonly called?

**Correct!**



Test set

*You are correct! The answer “*Test set*” is correct because in machine learning, the final estimate of outcome is done on test sets.*



Development set



Validation set



Training set

**Question 3**

**1 / 1 pts**

Consider the following Python code:

**diamond\_training\_data, diamond\_validation\_data, diamond\_test\_data = np.split(diamond\_data, [1500, 1800])**

For the datasets of 2000 rows, what would be the size of the Diamond\_test\_data?

**Correct!**



200

*You are correct! The answer “*200*” is correct because in the given code, the initial division point is 1500 and the second is 1800; therefore, the remaining 200 rows out of 2000 will be the size of*diamond\_test\_data*.*



1500



1800

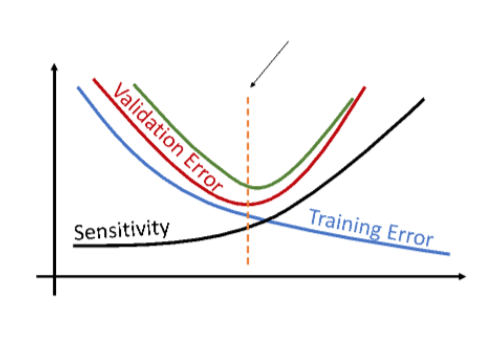


500

**Question 4**

**1 / 1 pts**

In the given graph, the green curve represents development error.





True

**Correct!**



False

*You are correct! The answer “*False*” is correct because* *the green curve in the graph represents test error.*