Question 1

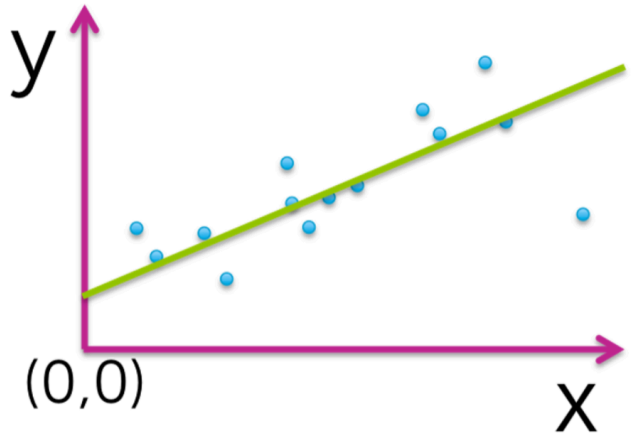
The following table illustrates the results of evaluating 4 models with different parameter choices on some data set. Which of the following models fit this data the best?

Model index	Parameters (intercept, slope)	Residual sum of squares (RSS)
1	(0,1.4)	20.51
2	(3.1,1.4)	15.23
3	(2.7, 1.9)	13.67
4	(0, 2.3)	18.99

a. Model 3
b. Model 1
c. Model 2
d. Model 4

Question 2

Assume we fit the following quadratic function: f(x) = w0+w1*x+w2*(x^2) to the dataset shown (blue circles). The fitted function is shown by the green curve in the picture below. Out of the 3 parameters of the fitted function (w0, w1, w2), which ones are estimated to be 0? *(Note: you must select all parameters estimated as 0 to get the question correct.)*



a. w2
b. w1
c. None of the above
d. w0

Question 3

Which of the following is **NOT** a **linear** regression model. *Hint: remember that a linear regression model is always linear in the parameters, but may use non-linear features.*

a. y = w0 + w1x
b. y = w0 + w1log(x)
c. y = w0w1 + log(w1)x

d. $y = w0 + w1x2$

Question 4
Gradient descent/ascent is…

a. An approximation to simple linear regression
b. An algorithm for minimizing/maximizing a function
c. A model for predicting a continuous variable
d. A theoretical statistical result
e. A modeling technique in machine learning

Question 5
Gradient descent/ascent allows us to…

a. Predict a value based on a fitted function
b. Assess performance of a model on test data
c. Estimate model parameters from data

Question 6
Which of the following statements about learning rate in gradient descent is/are **TRUE** (select all that apply)

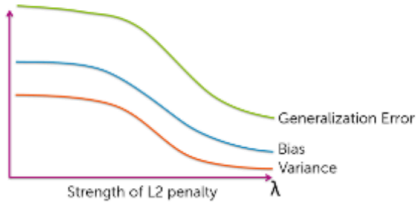
a. If the learning rate is too small (but not zero) gradient descent may take a very long time to converge
b. It's important to choose a very small learning rate
c. The learning rate doesn't matter
d. If the learning rate is too large gradient descent may not converge

Question 7
In ridge regression, choosing a large penalty strength λ tends to lead to a model with (choose all that apply):

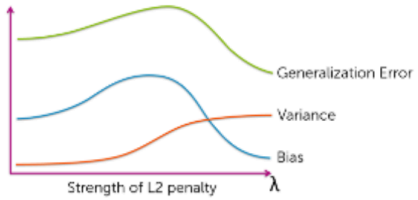
a. High variance
b. High bias
c. Low bias
d. Low variance

Question 8
Which of the following plots best characterize the trend of bias, variance, and generalization error (all plotted over λ)?

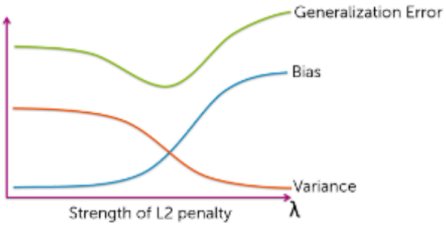
A:



B:



C:



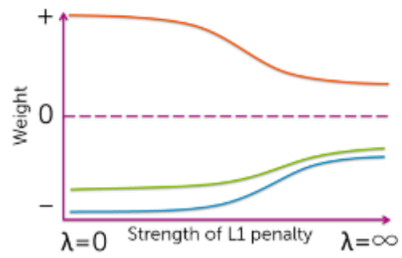
a. C
b. A
c. B

Question 9

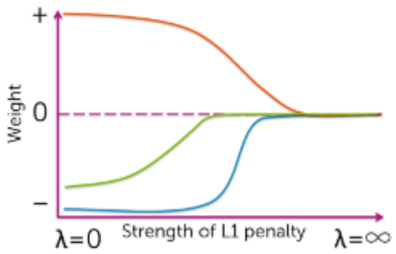
Which of the plots could correspond to a lasso coefficient path?

Hint: notice λ = Infinity in the bottom right of the plots. How should coefficients behave eventually as λ goes to infinity?

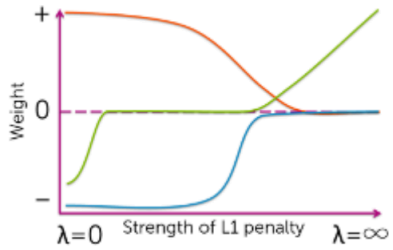
A



B



C



a. B
b. A
c. C

Question 10

Each time we update the parameters using gradient descent, we obtain parameters that will make the loss smaller.

a. True
b. False