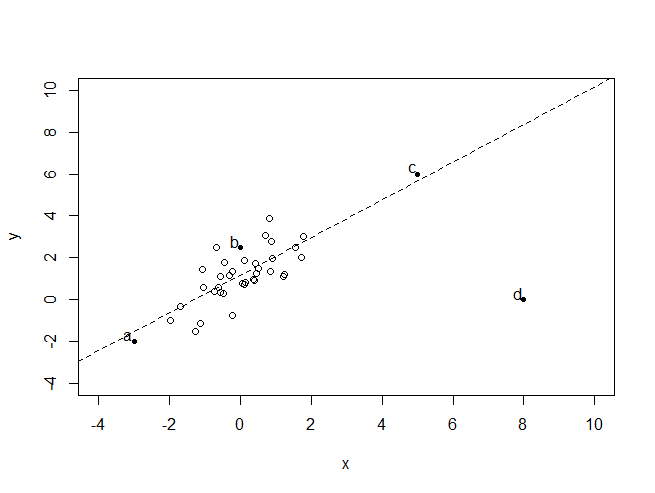
## Week 1 – Simple Linear Regression

1. Assume you fit a regression model to predict house prices from square feet based on a training data set consisting of houses with square feet in the range of 1000 to 2000. In which interval would we expect predictions to do best?
   1. [1000, 2000]
2. In a simple regression model, if you increase the input value by 1, then you expect the output to change by:
   1. The value of the *slope* parameter.
3. Two people present you with fits of their simple regression model for predicting house prices from square feet. You discover that the estimated intercept and slopes are exactly the same. This necessarily implies that these two people fit their models on exactly the same data set.
   1. False
4. You have a dataset consisting of the sales prices of houses in your neighbourhood, with each sale time-stamped by the month and year in which the house sold. You want to predict the average value of houses in your neighbourhood over time, so you fit a simple regression model with average house price as the output and the time index (in months) as the input. Based on 10 months of data, the estimated intercept is $4,569 and the estimated slope is 143 ($/month). If you extrapolate this trend forward in time, at which time index (in months) do you predict that your neighbourhood’s value will have doubled relative to the value at month (index) 10? (Assume months are 0-indexed).
   1. 52
5. Your friend in the US gives you a simple regression fit for predicting house prices from square feet. The estimated intercept is -44,850 and the estimated slop is 280.76. You believe that your housing market behaves very similarly, but houses are measured in square metres. To make predictions for inputs in square metres, what intercept must you use?
   1. -44,850
6. To make predictions for inputs in square meters, what slope must you use?
7. Consider the following dataset and the regression line fitted on this data:



Which bold/labelled point, if removed, will have the largest effect on the fitted regression line?

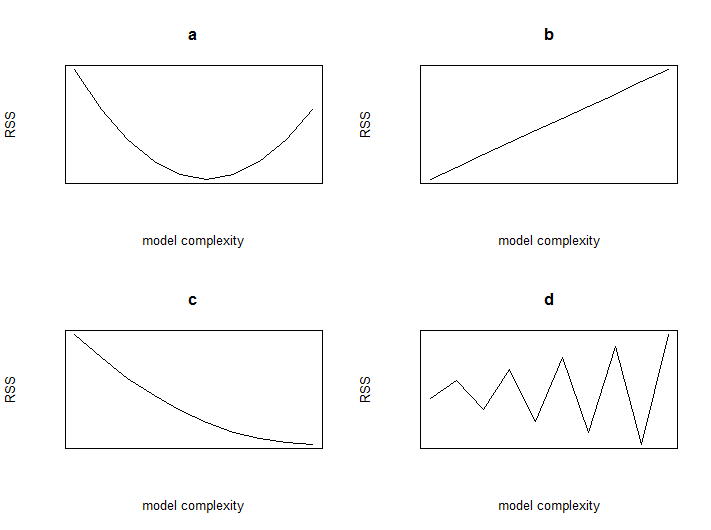
* 1. D.

## Week 2 – Multiple Regression

1. Which of the following is NOT a linear regression model.
2. Your estimated model for predicting house prices has a large positive weight on 'square feet living'. This implies that if we remove the feature 'square feet living' and refit the model, the new predictive performance will be worse than before.
   1. False.
3. Your estimated model for predicting house prices has a positive weight on 'square feet living'. You then add 'lot size' to the model and re-estimate the feature weights. The new weight on 'square feet living' [\_\_\_\_\_\_\_\_\_] be positive.
   1. Might.
4. If you double the value of a given feature (i.e. a specific column of the feature matrix), what happens to the least-squares estimated coefficients for every other feature? (Assume you have no other feature that depends on the doubled feature i.e. no interaction terms).
   1. They stay the same.
5. Gradient descent/ascent is:
   1. An algorithm for maximising/minimising a function.
6. Gradient descent/ascent allows us to:
   1. Estimate model parameters from data.
7. Which of the following statements about step-size in gradient descent are true?
   1. If the step-size is too large, gradient descent may not converge.
   2. If the step-size is too small (but not zero), gradient descent may take a very long time to converge.
8. Let's analyse how many computations are required to fit a multiple linear regression model using the closed-form solution based on a data set with 50 observations and 10 features. In the videos, we said that computing the inverse of the matrix was on the order of operations. Let's focus on forming this matrix prior to inversion. How many multiplications are required to form the matrix ?
9. More generally, if you have features and observations what is the total complexity of computing ?

## Week 3 – Assessing Performance

1. If the features of Model 1 are a strict subset of Model 2, the training error of the two models can never be the same.
   1. False.
2. If the features of Model 1 are a strict subset of Model 2, which model will usually have the lowest training error?
   1. Model 2.
3. If the feature of Model 1 are a strict subset of Model 2, which model will usually have the lowest test error?
   1. It’s impossible to tell with only this information.
4. If the features of Model 1 are a strict subset of those in Model 2, which model will usually have lower bias?
   1. Model 2.
5. Which of the following plots of model complexity vs RSS is most likely from the training data?



* 1. C

1. Which of the plots of model complexity vs RSS is most likely from the test data?
   1. A
2. It is always optimal to add more features to a regression model?
   1. False
3. A simple model with few parameters is most likely to suffer from:
   1. High bias.
4. A complex model with many parameters is most likely to suffer from:
   1. High variance.
5. A model with many parameters that fits training data very well, but does poorly on test data is considered to be:
   1. Overfitted.
6. A common process for selecting a parameter like the optimal polynomial degree is:
   1. Minimising validation error.
7. Selecting model complexity on test data:
   1. Provides an overly optimistic assessment of performance of the resulting model.
   2. Should never be done.
8. For a fixed model complexity, in the limit of an infinite amount of training data:
   1. Variance goes to 0.

## Week 4 – Ridge Regression