CS150 Quiz11

* Required

K-Means

2. **1) K-means is guaranteed to converge within X iterations, where X equals: ***
   Let N be size of data, K is number of centers
   Mark only one oval.

   - 1
   - 10
   - K
   - N
   - None of the above

3. **2) The best value of K should always be at least 10. ***
   Mark only one oval.

   - True
   - False
   - Not enough information

4. **3) K Means++ is K-Means with a different way of initializing the centers. ***
   Mark only one oval.

   - True
   - False
   - Not enough information

K-Means

5. **4) Suppose we've already run K-means with k=3, and have the following cluster centers: {Red=(1, 6), Green=(5, 3), Blue=(2, 2)}. We then receive a new point (4, 5), which cluster do you predict it belongs to?** *
   Mark only one oval.

   - ◯ Red
   - ◯ Green
   - ◯ Blue
   - ◯ None of the above

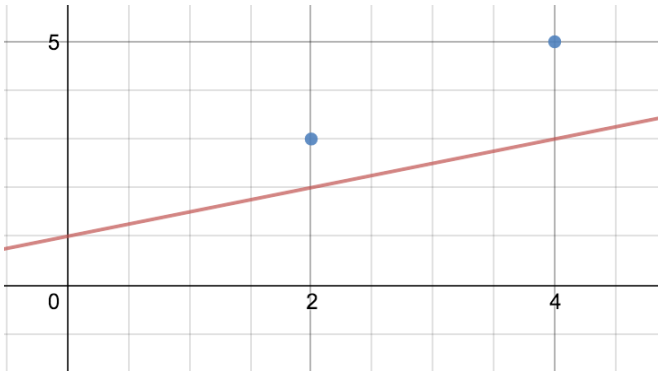
6. **5) The Reservoir Sampling algorithm has an approximate runtime of:** *
   K = number of centers, N = size of dataset, d = number of dimensions
   Mark only one oval.

   - ◯ O(1)
   - ◯ O(Kd)
   - ◯ O(N)
   - ◯ O(N**2)

7. **6) K-Means++ initialization will not work if our dataset is too large to fit in memory** *
   Mark only one oval.

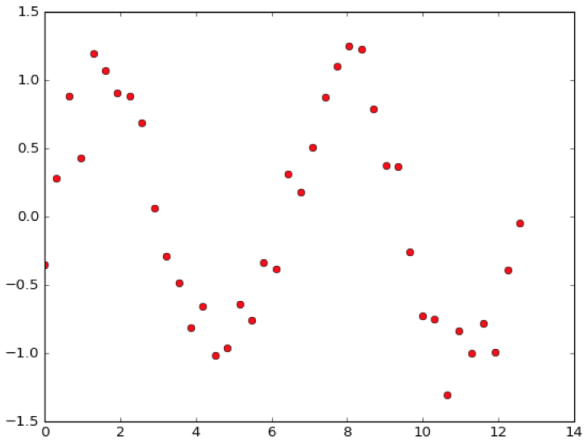
   - ◯ True
   - ◯ False because we don't need all of the data
   - ◯ False because we can use a streaming algorithm to sample our initial points
   - ◯ Not enough information

Linear Regression



8. **7) We have two data points (2, 3), (4, 5) and we've computed our regression line as y = ½x + 1. What is our mean squared error?** *
   Write fractional answers as a decimal number.

   _____



9. **8) Suppose we wanted to apply linear regression to this data. Which of the following features would you include to better fit the data?** *
   Check all that apply.

   - [ ] 1
   - [ ] x
   - [ ] x^2
   - [ ] x^3
   - [ ] x!
   - [ ] Not possible to apply linear regression to this data

10. **9) Increasing the number of features will guarantee your model to perform better.** *
    Mark only one oval.

    - ◯ True because you have more information, and more is always better
    - ◯ False because your computational performance will slow drastically
    - ◯ False because your model may overfit on the training data

11. **10) Mark the following true statements regarding regularization:** *
    Assume bias refers to squared bias
    Check all that apply.

    - [ ] Regularization is a way of mitigating overfitting
    - [ ] Increasing the regularization parameter will decrease bias
    - [ ] Increasing the regularization parameter will increase bias
    - [ ] Decreasing the regularization parameter will increase variance
    - [ ] Decreasing the regularization parameter will decrease variance

Make sure you check the box below.