CS639 – Final Practice Questions

Instructions:

* These questions are similar to those in the final exam.
* This document is not representative of the length of the final (it’s too long).
* Go back and study the material that is associated with each of the practice questions.
* For additional exam-samples please take a look here: <http://www.ds100.org/sp19/resources>

# **Data Generation and Probability Samples**

For each of the following questions select the **single best answer.**

1. A political scientist is interested in answering a question about a country composed of three states with exactly 10000, 20000, and 30000 voting adults. To answer this question, a political survey is administered by randomly sampling 25, 50, and 75 voting adults from each town, respectively. Which sampling plan was used in the survey?

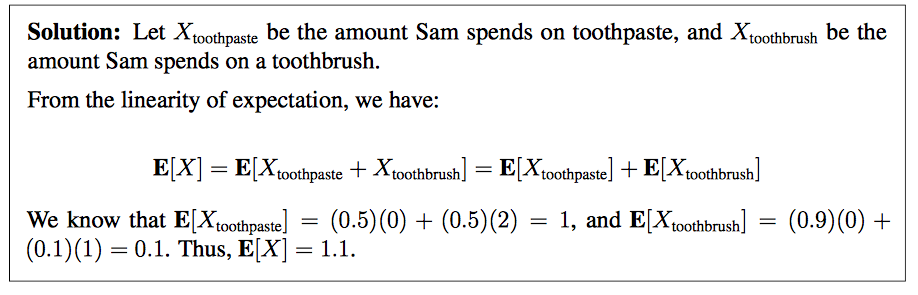
* Cluster sampling
* Stratified Sampling
* Quota sampling
* Snowball sampling

2. A deck with 26 cards labeled A through Z is thoroughly shuffled, and the value of the **third** card in the deck is recorded. What is the probability that we observe the letter C on the third card?

* 1/26
* 3/26
* (25/26) \* (24/26) \* (1/26)
* None of the above.

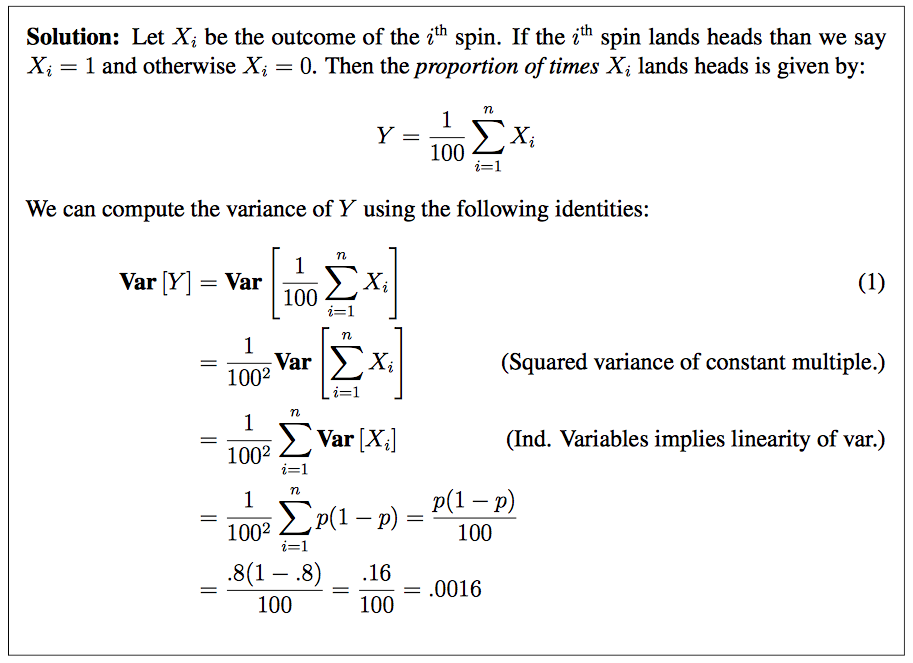
3. Suppose Iris visits your store to buy some items. She buys toothpaste for $2.00 with probability 0.5. She buys a toothbrush for $1.00 with probability 0.1. Let the random variable X be the total amount Sam spends. What is E[X]? Show your work in the space provided.

* $1.10
* $1.5
* $3.00
* The toothpaste purchase may not be independent of the toothbrush purchase so we cannot compute this expectation.



4. Suppose we have a coin that lands heads 80% of the time. Let the random variable X be the *proportion* of times the coin lands tails out of 100 flips. What is Var[X]? You must show your work in the space provided.

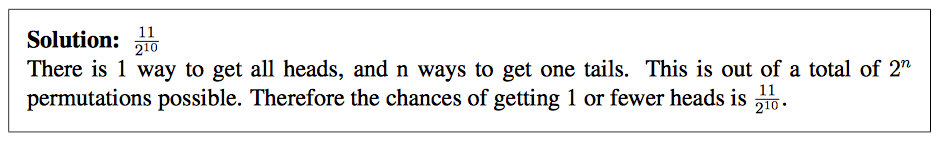
* 0.8
* 0.16
* 0.04
* 0.0016
* 0.008



# Hypothesis Testing

A mysterious undergraduate stops you on your way to class and claims that they have learned to flip any coin such that it lands on heads more often than the 50% you’d expect from random chance. To demonstrate, they takes a penny from their wallet, flip it 10 times, and get heads nine times and only gets tails once.

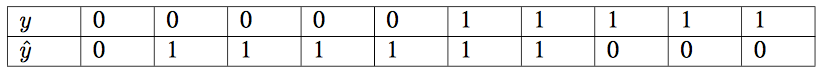
1. The null hypothesis is that this was pure random chance, and that the probability of getting heads was 50% for each flip. What is the p-value under the null hypothesis of getting 1 or fewer tails out of 10 flips?



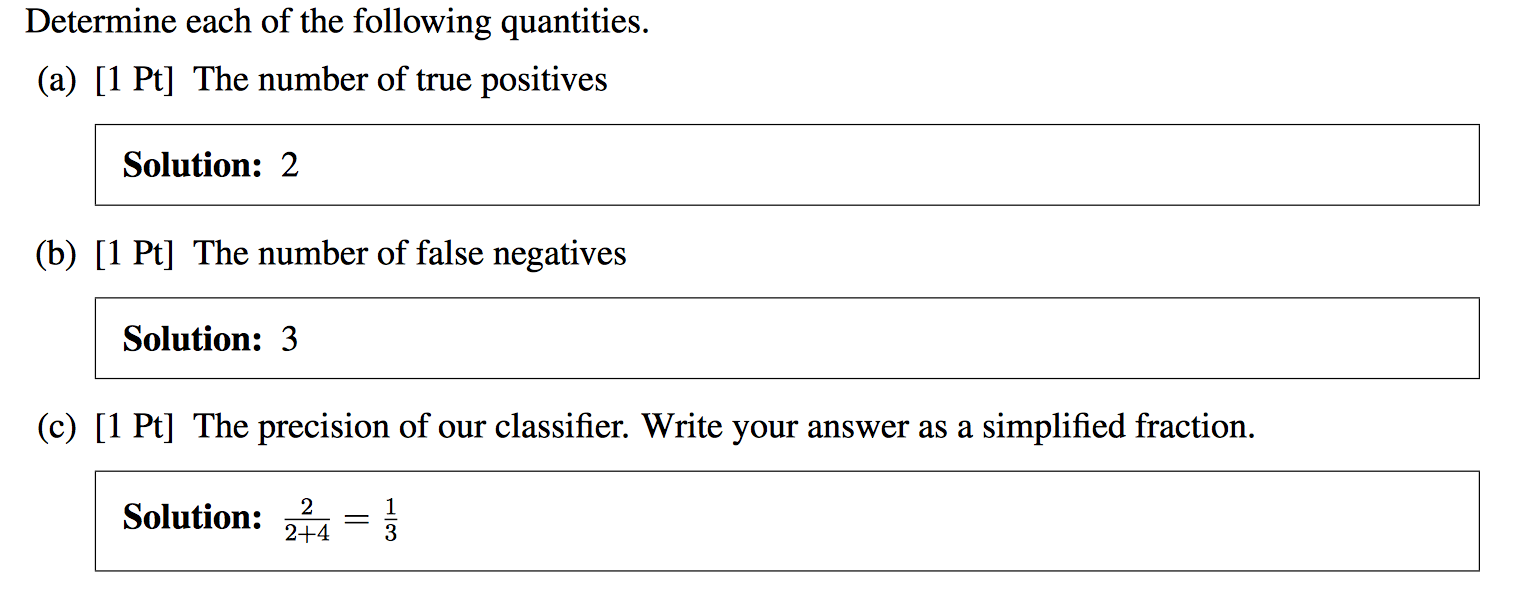
1. Suppose the undergraduate flips the coin 28 more times, and they all end up heads. The resulting p value including all 38 flips under the null hypothesis is approximately pb = 10-10. Which of the following are true? Select all that apply,
   * It is extremely unlikely that the undergraduate just happened to get 37 heads by randomly getting heads on 50/50 coin flips.
   * pb is the probability that the null hypothesis is true.
   * 1 - pb is the probability that the undergraduate has the skill to flip any arbitrary coin and get heads.
   * If you flipped a fair coin 38 times, pb is the chance that you’d get at least 37 heads by random chance.
   * The undergraduate has proven beyond any reasonable doubt that she has the skill to flip any coin to land on heads with high probability.
   * None of the above.

# Classification

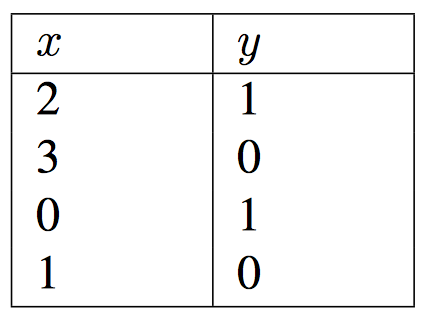
1. Suppose we train a binary classifier on some dataset. Suppose y is the set of true labels, and yˆ is the set of predicted labels.

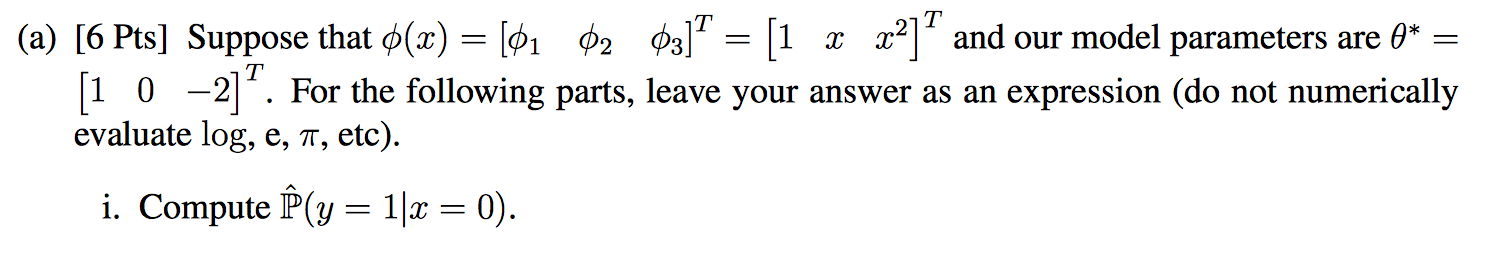


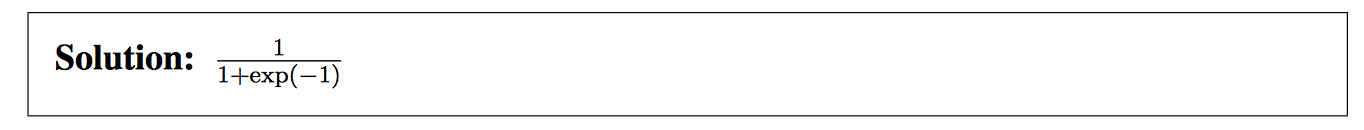
Determine each of the following quantities

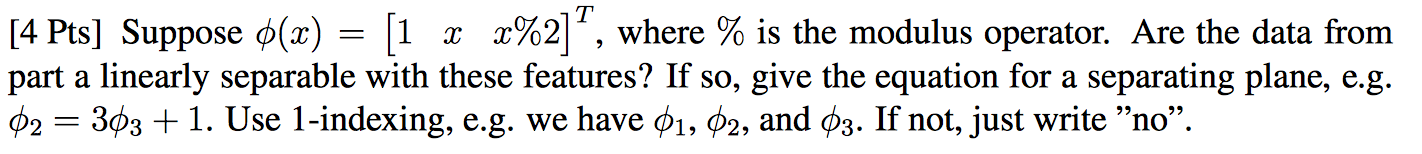


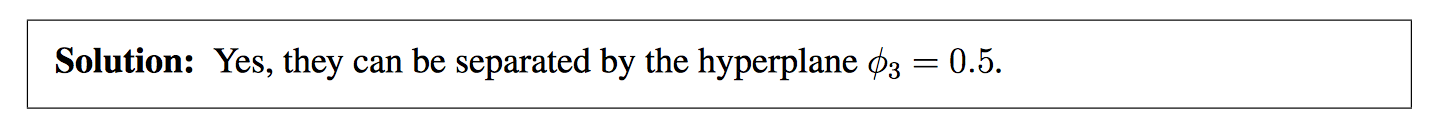
1. You have a classification data set, where *x* is some value and y is the label for that value:

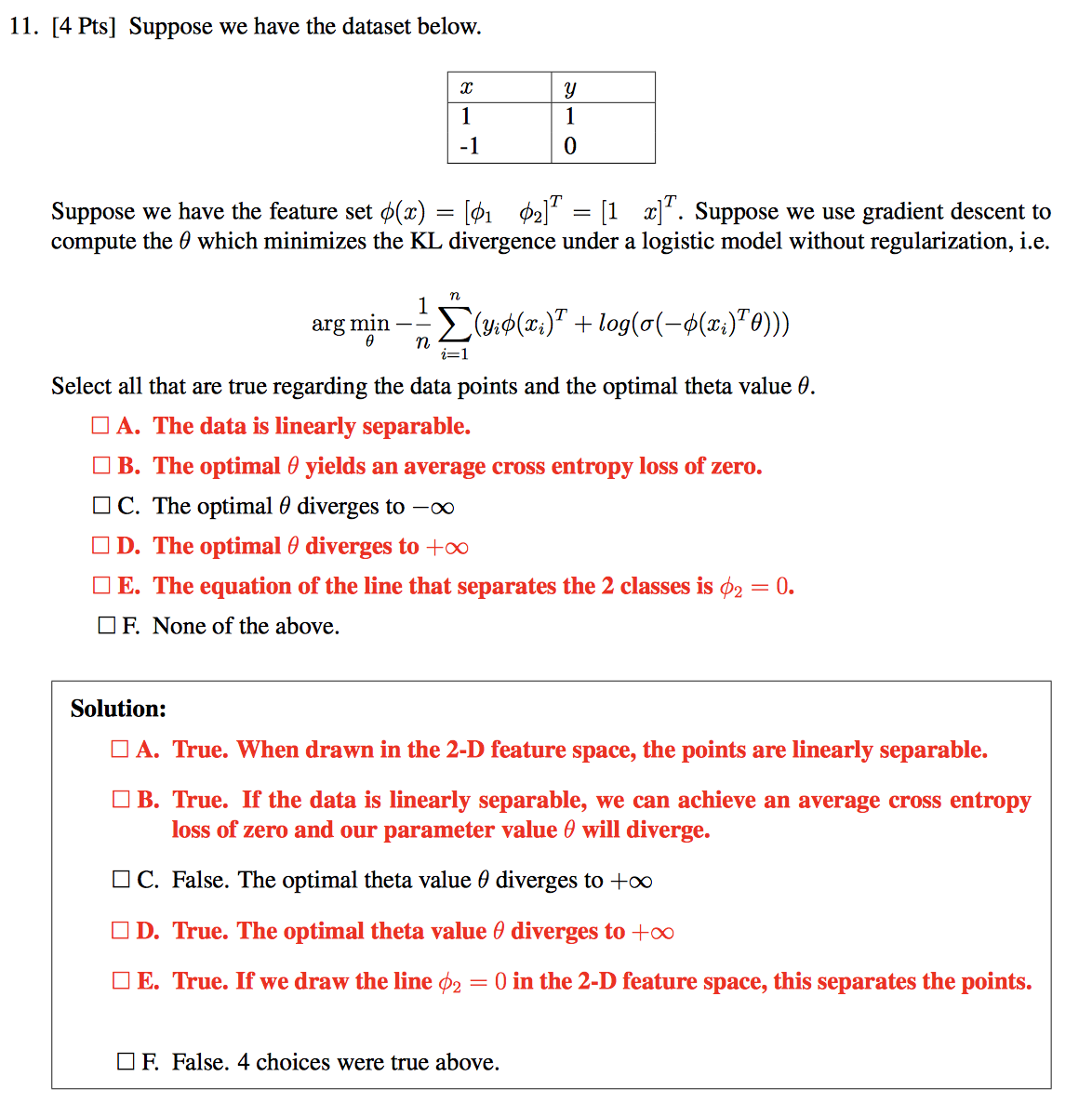


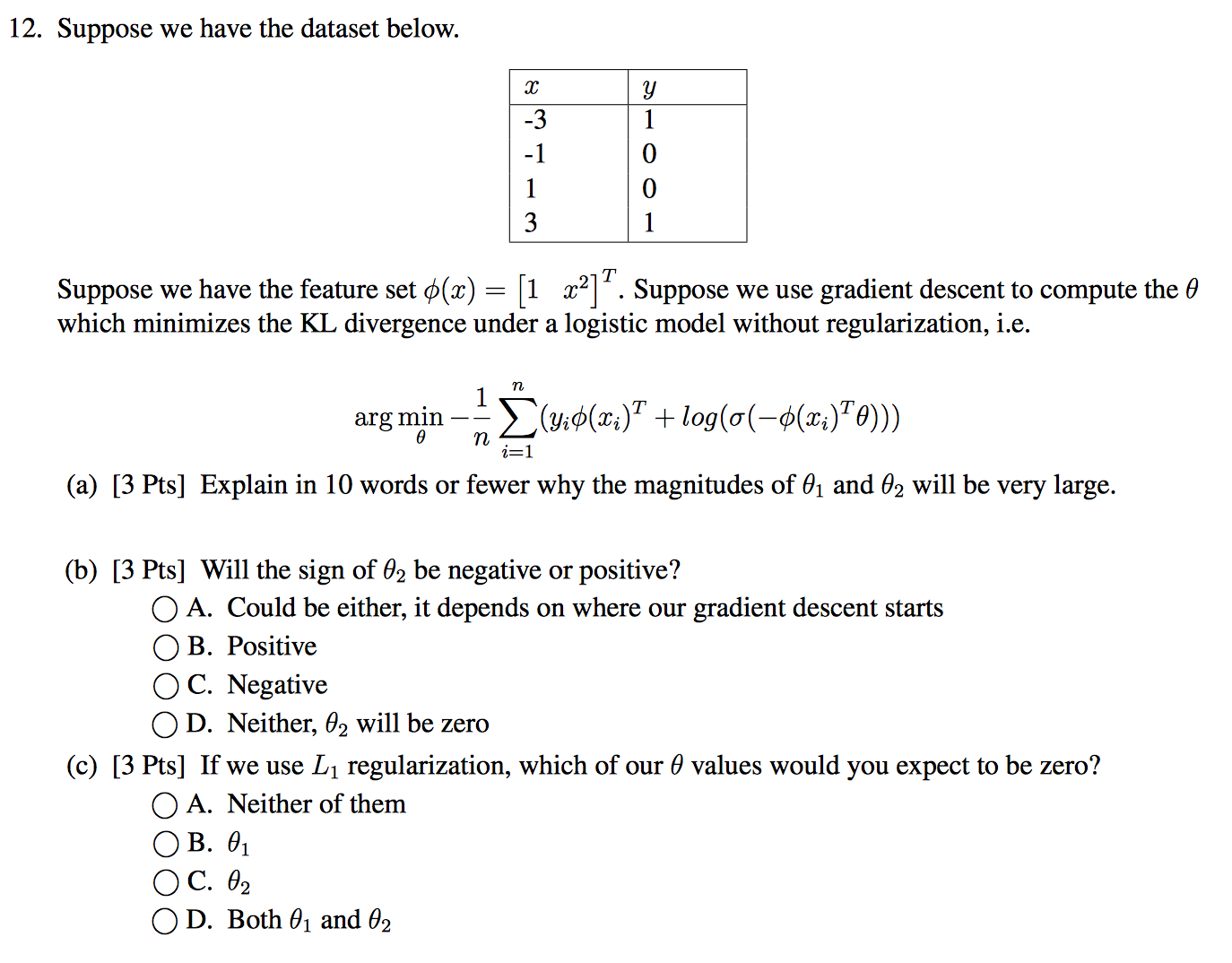


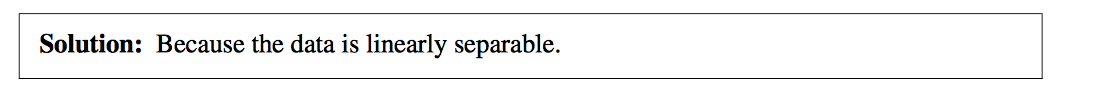




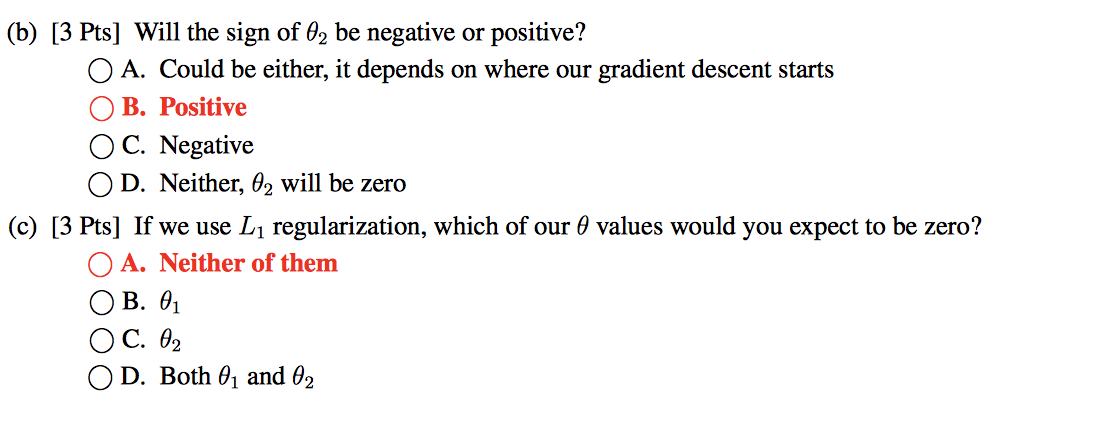




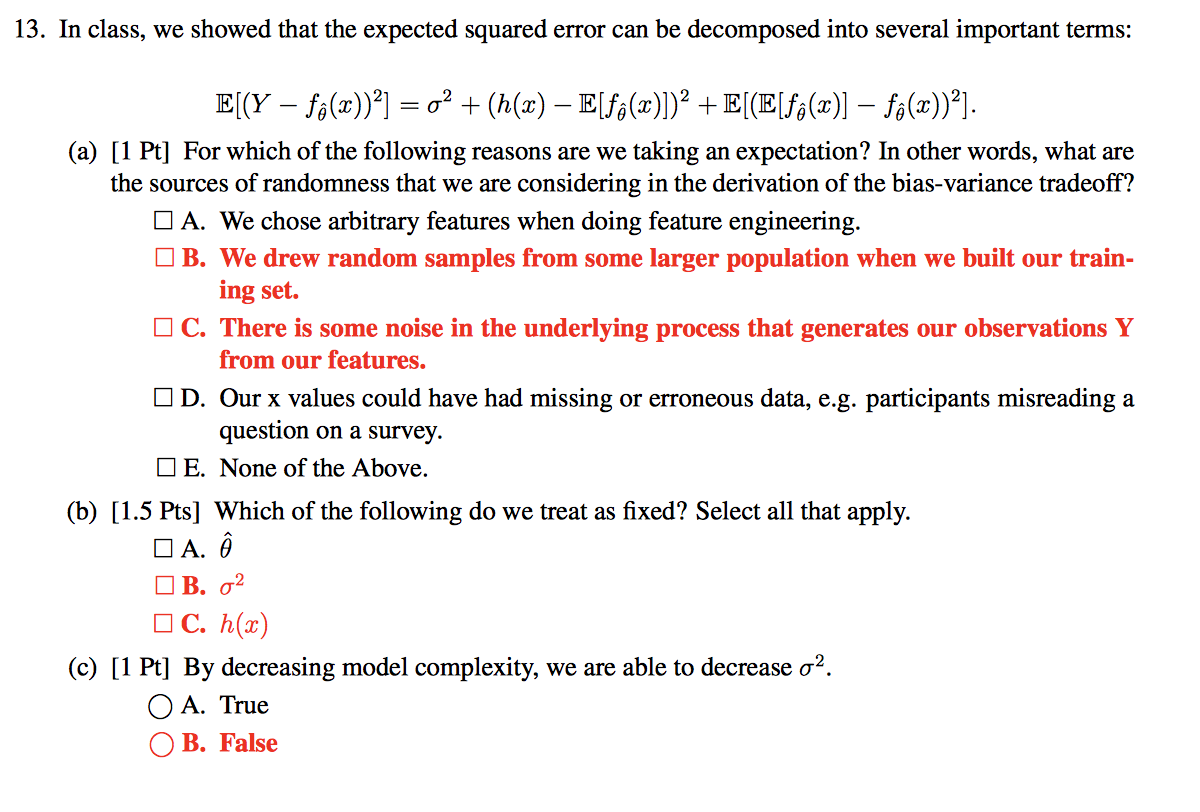


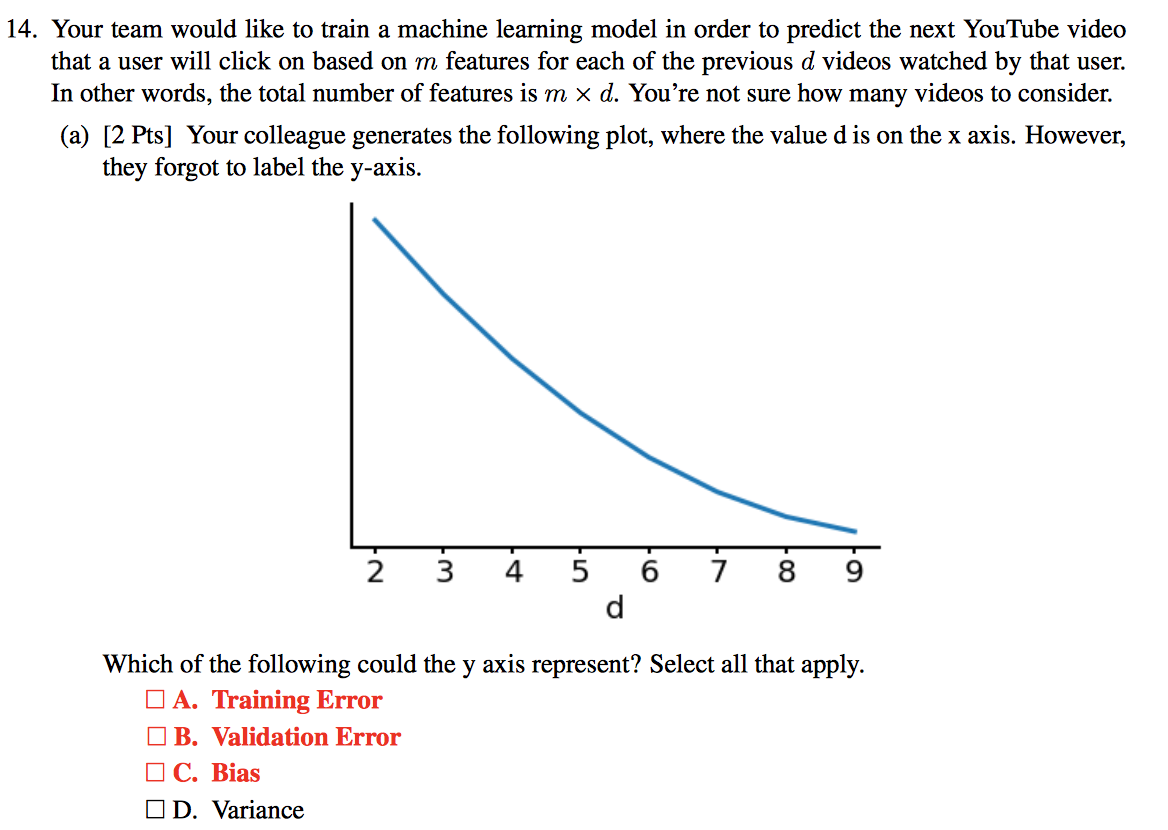
Solution to (a) 

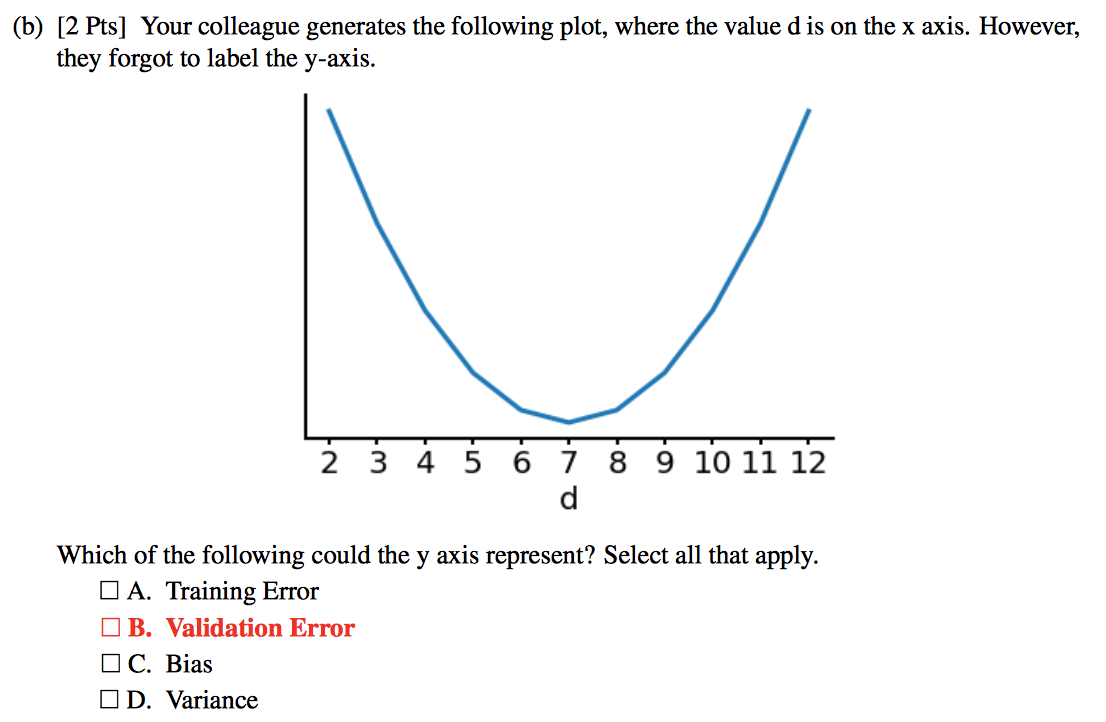
Solution to (b) and (c)



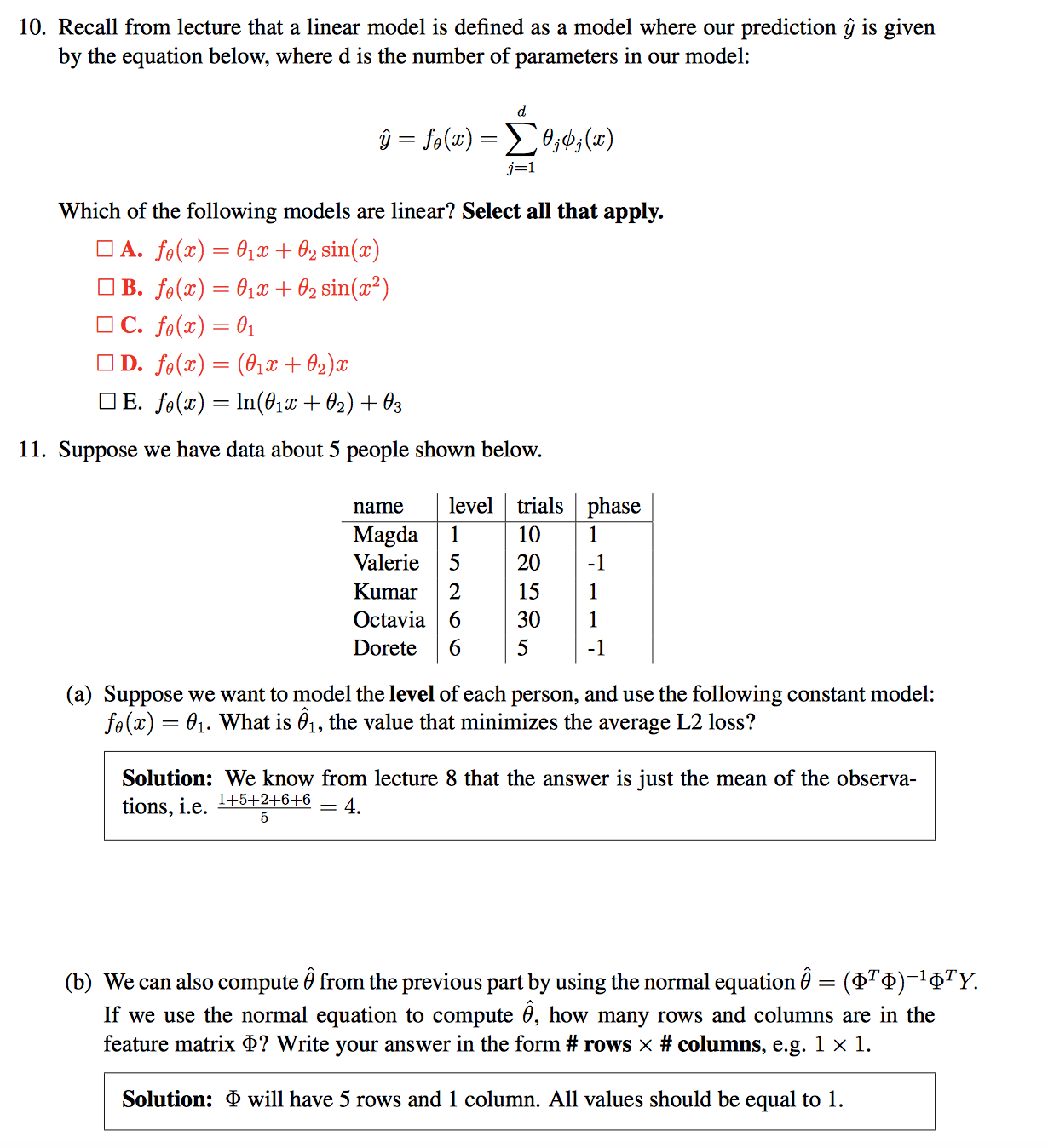
# Bias and Variance Tradeoff

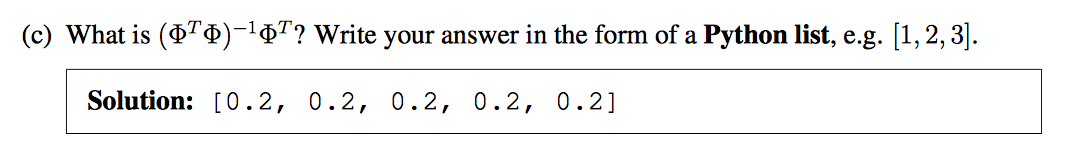




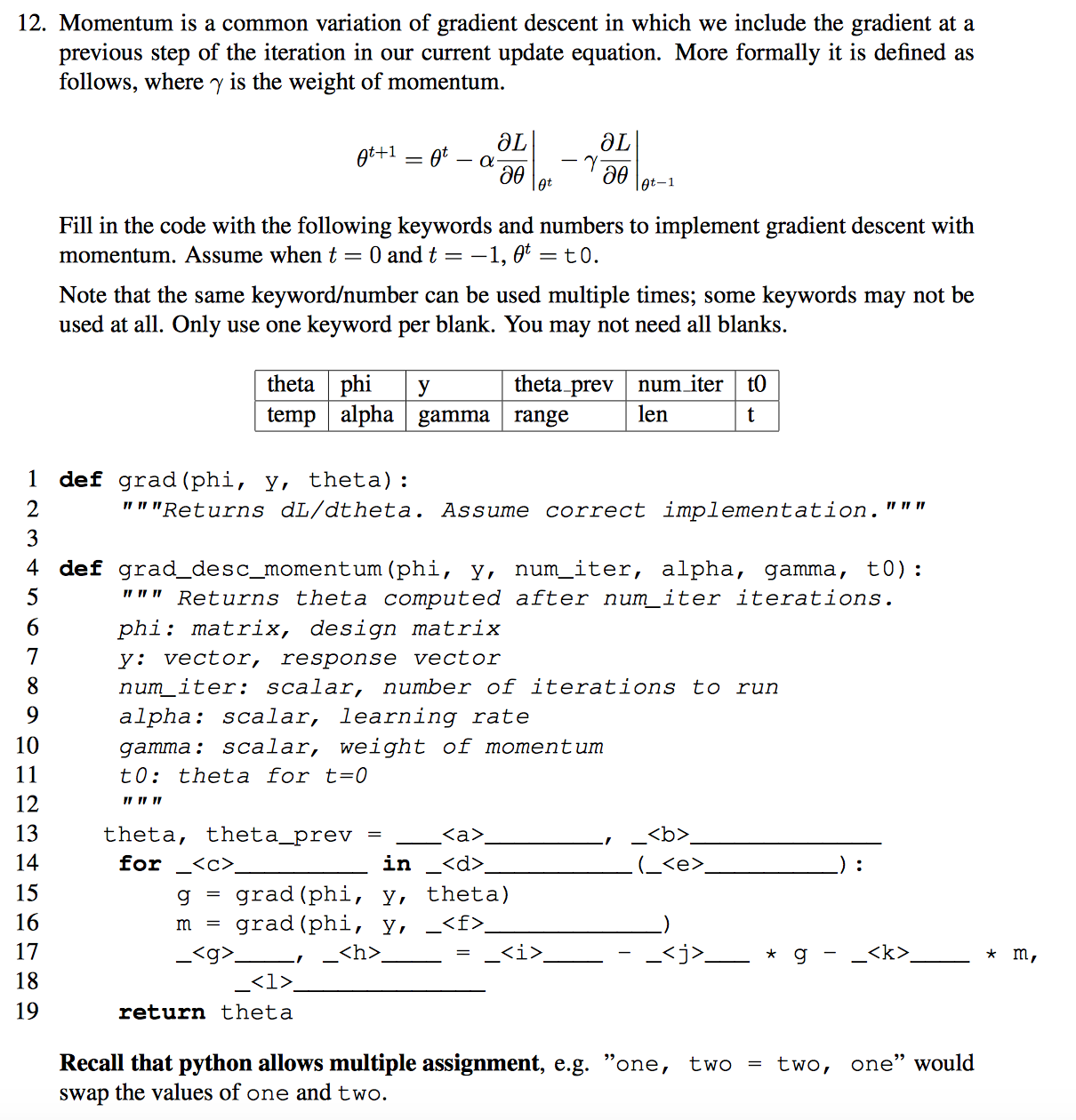


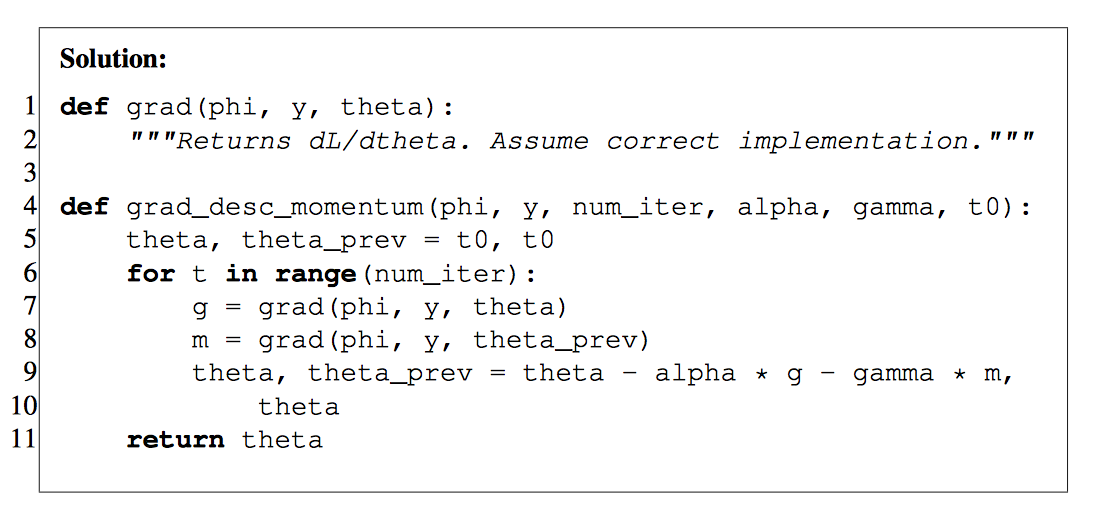
# Linear Models





# Gradient Descent





# Gradient Descent

