1 point

1. 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.*

- ( ) $y = w_0 + w_1 x$
- ( ) $y = w_0 + w_1 x^2$
- ( ) $y = w_0 + w_1 \log(x)$
- (•) $y = w_0 w_1 + \log(w_1) x$

---

1 point

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.

- ( ) True
- (•) False

---

1 point

3. *Complete the following:* 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.

- ( ) will not
- ( ) will definitely
- (•) might

---

1 point

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).

- ( ) They double
- ( ) They halve
- (•) They stay the same
- ( ) It is impossible to tell from the information provided

---

1 point

5. Gradient descent/ascent is...

- ( ) A model for predicting a continuous variable
- (•) An algorithm for minimizing/maximizing a function
- ( ) A theoretical statistical result
- ( ) An approximation to simple linear regression
- ( ) A modeling technique in machine learning

---

1 point

6. Gradient descent/ascent allows us to...

- ( ) Predict a value based on a fitted function
- (•) Estimate model parameters from data
- ( ) Assess performance of a model on test data

---

1 point

7. Which of the following statements about step-size in gradient descent is/are **TRUE** (select all that apply)

- [ ] It's important to choose a very small step-size
- [ ] The step-size doesn't matter
- [x] If the step-size is too large gradient descent may not converge
- [x] If the step size is too small (but not zero) gradient descent may take a very long time to converge

---

1 point

8. Let's analyze 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 10x10 matrix $H^T H$ was on the order of D^3 operations. Let's focus on forming this matrix **prior** to inversion. How many multiplications are required to form the matrix $H^T H$?

Please enter a number below.

5000

---

1 point

9. More generally, if you have D features and N observations what is the total complexity of computing $(H^T H)^{-1}$?

- ( ) $O(D^3)$
- ( ) $O(ND^3)$
- (•) $O(ND^2 + D^3)$
- ( ) $O(ND^2)$
- ( ) $O(N^2 D + D^3)$
- ( ) $O(N^2 D)$