

Name: \_\_\_\_\_  
Reading Quiz 1

Give the best answer(s) among those provided. Some questions have more than one answer.

- 1) What is multivariate regression?

- a. Single input, multiple outputs
- b. Multiple inputs, multiple outputs
- c. Multiple inputs, one output

**Simple regression** pertains to **one** dependent variable ( $y$ ) and **one** independent variable ( $x$ ):  
 $y = f(x)$

**Multiple regression (aka multivariable regression)** pertains to **one** dependent variable and **multiple** independent variables:  $y = f(x_1, x_2, \dots, x_n)$

**Multivariate regression** pertains to **multiple** dependent variables and **multiple** independent variables:  $y_1, y_2, \dots, y_m = f(x_1, x_2, \dots, x_n)$ . You may encounter problems where both the dependent and independent variables are arranged as matrices of variables (e.g.  $y_{11}, y_{12}, \dots$  and  $x_{11}, x_{12}, \dots$ ), so the expression may be written as  $Y = f(X)$ , where capital letters indicate matrices.

- 2) You design a machine-learning model to study a problem. How do you know that your model has improved your knowledge?

- a. You get the answer you expect
- b. You divide your data into a test set and a train set, and the performance on your training set is better than the performance on your test set.
- c. Your performance is significantly better than random guessing

- 3) How is precision in binary classification defined?

- a.  $(\text{True Positives}) / (\text{True Positives} + \text{False Positives})$
- b.  $(\text{True Positives}) / (\text{True Positives} + \text{False Negatives})$
- c. The percentage of identified signal samples that were correctly classified.
- d. The percentage of true signal samples that were correctly classified.

- 4) How is recall in binary classification defined?

- a.  $(\text{True Positives}) / (\text{True Positives} + \text{False Positives})$
- b.  $(\text{True Positives}) / (\text{True Positives} + \text{False Negatives})$
- c. The percentage of identified signal samples that were correctly classified.
- d. The percentage of true signal samples that were correctly classified.

- 5) Which of the following are true about a ROC curve?

- a. The closer the curve is to the diagonal (from lower left to upper right) the better the model
- b. You can use it to compare two different models for a single binary classification problem
- c. You can use it to determine the performance of your model at a variety of different thresholds.
- d. You only need one performance point at a single threshold to calculate it.

- 6) Which is generally better more helpful in evaluating your model's performance?

- a. The full ROC curve.
- b. AUC: that single number has all of the information you will need!