Data 100, Spring 2022	**February 25, 2022**

Discussion #6

Linear Models

1. Which of the following models are linear? Select all that apply.
   - ☐ A. $\hat{y} = \theta_1 x + \theta_2 \sin(x)$
   - ☐ B. $\hat{y} = \theta_1 x + \theta_2 \sin(x^2)$
   - ☐ C. $\hat{y} = \theta_1$
   - ☐ D. $\hat{y} = (\theta_1 x + \theta_2)\, x$
   - ☐ E. $\hat{y} = \ln(\theta_1 x + \theta_2) + \theta_3$

2. Which of the following are true about the optimal solution $\hat{\theta}$ to ordinary least squares (OLS)? Recall that the least squares estimate $\hat{\theta}$ solves the normal equation $(\mathbb{X}^T\mathbb{X})\theta = \mathbb{X}^T\mathbb{Y}$.

$$\hat{\theta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y}$$

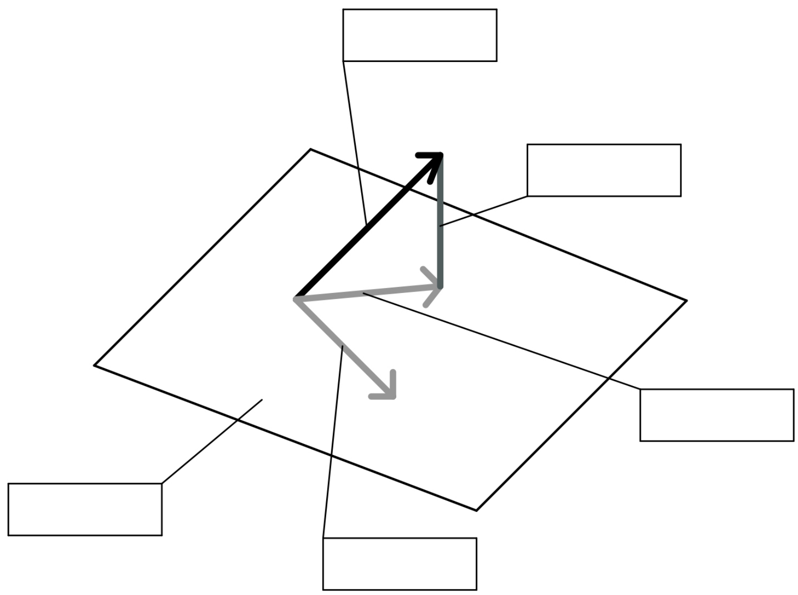
   - ☐ A. Using the normal equation, we can derive an optimal solution for simple linear regression with an L_2 loss.
   - ☐ B. Using the normal equation, we can derive an optimal solution for simple linear regression with an L_1 loss.
   - ☐ C. Using the normal equation, we can derive an optimal solution for a constant model with an L_2 loss.
   - ☐ D. Using the normal equation, we can derive an optimal solution for a constant model with an L_1 loss.
   - ☐ E. Using the normal equation, we can derive an optimal solution for the model specified option B in question 1 ($\hat{y} = \theta_1 x + \theta_2 \sin(x^2)$).

3. Which of the following conditions are required for the least squares estimate in Question 2?
   - ☐ A. $\mathbb{X}$ must be full column rank.
   - ☐ B. $\mathbb{Y}$ must be full column rank.
   - ☐ C. $\mathbb{X}$ must be invertible.
   - ☐ D. $\mathbb{X}^{\mathbb{T}}$ must be invertible.

Geometry of Least Squares

4. Suppose we have a dataset represented with the design matrix span($\mathbb{X}$) and response vector $\mathbb{Y}$. We use linear regression to solve for this and obtain optimal weights as $\hat{\theta}$. Label the following terms on the geometric interpretation of ordinary least squares:

   - $\mathbb{X}$ (i.e., span($\mathbb{X}$))
   - The response vector $\mathbb{Y}$
   - The residual vector $\mathbb{Y} - \mathbb{X}\hat{\theta}$
   - The prediction vector $\mathbb{X}\hat{\theta}$ (using optimal parameters)
   - A prediction vector $\mathbb{X}\alpha$ (using an arbitrary vector α).



   (a) What is always true about the residuals in least squares regression? Select all that apply.

   ☐ A. They are orthogonal to the column space of the design matrix.
   ☐ B. They represent the errors of the predictions.
   ☐ C. Their sum is equal to the mean squared error.
   ☐ D. Their sum is equal to zero.
   ☐ E. None of the above.

   (b) Which are true about the predictions made by OLS? Select all that apply.

   ☐ A. They are projections of the observations onto the column space of the design matrix.
   ☐ B. They are linear combinations of the features.
   ☐ C. They are orthogonal to the residuals.
   ☐ D. They are orthogonal to the column space of the features.

☐ E. None of the above.

(c) We fit a simple linear regression to our data $(x_i, y_i), i = 1, 2, 3$, where x_i is the independent variable and y_i is the dependent variable. Our regression line is of the form $\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$. Suppose we plot the relationship between the residuals of the model and the $\hat{y}s$, and find that there is a curve. What does this tell us about our model?

☐ A. The relationship between our dependent and independent variables is well represented by a line.

☐ B. The accuracy of the regression line varies with the size of the dependent variable.

☐ C. The variables need to be transformed, or additional independent variables are needed.

(d) Which are the following is true of the mystery quantity $\vec{v} = (I - \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T)\mathbb{Y}$?

☐ A. The vector $\vec{v}$ represents the residuals for any linear model.

☐ B. If the $\mathbb{X}$ matrix contains the $\vec{1}$ vector, then the sum of the elements in vector $\vec{v}$ is 0 (i.e. $\sum_i v_i = 0$).

☐ C. All the column vectors x_i of $\mathbb{X}$ are orthogonal to $\vec{v}$.

☐ D. If $\mathbb{X}$ is of shape n by p, there are p elements in vector $\vec{v}$.

☐ E. For any α, $\mathbb{X}\alpha$ is orthogonal to $\vec{v}$.

Linear Regression Fundamentals (Extra)

5. In this problem, we will review some of the core concepts in linear regression.

   Suppose we create a linear model with parameters $\hat{\theta} = [\hat{\theta}_0, \ldots, \hat{\theta}_p]$. As we saw in lecture, given an observation $\vec{x}$, such a model makes predictions $\hat{y} = \hat{\theta} \cdot \vec{x} = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 + \ldots + \hat{\theta}_p x_p$.

   (a) Suppose $\hat{\theta} = [2, 0, 1]$ and we receive an observation $\vec{x_1} = [1, 2, 3]$. What $\hat{y_1}$ value will this model predict for the given observation?

   (b) Suppose the true y_1 was 3.5. What will be the L_2 loss for our prediction $\hat{y_1}$ from the previous part?

   (c) Suppose we receive another observation $\vec{x_2} = [1, 5, 1]$. What $\hat{y_2}$ value will this model predict for the given observation?

   (d) Suppose the true y_2 was 4. What will be the mean squared error of our model, given the two observations?