Names: ____________________

NetIDs: ____________________

Discussion section: D ________ **Group number:** ________

MATH 257 - WORKSHEET 12

Suppose we collected the following data:

Sample i	x_i	y_i
1	1	1
2	2	3
3	3	6
4	3	7

(1) Let's start with a linear model for the y-values in terms of the x-values, say $y = \beta_0 + \beta_1 x$ for some parameters $\beta_0, \beta_1 \in \mathbb{R}$ to be determined so as to optimize fit. Let's organize the data as follows: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_4 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix}$, and $\mathbf{w} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ a tuple of all 1s. If the data fit the model perfectly, what is the relation among the tuples $\mathbf{w}, \mathbf{x}, \mathbf{y}$? Can you write this relation as a matrix equation involving the unknown tuple $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$?

If the model were a perfect fit, we could find parameters $\beta_0, \beta_1 \in \mathbb{R}$ such that

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_1 \\ \vdots \\ \beta_0 + \beta_1 x_4 \end{pmatrix} = \beta_0 \mathbf{w} + \beta_1 \mathbf{x} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \mathbf{w} & \mathbf{x} \end{pmatrix} \beta$$

(2) The matrix is called a *model matrix* (often denoted X), $\mathbf{y}$ the *response vector*, and β the *parameter vector*. Since the data is not expected to fit the model perfectly, the preceding matrix equation is generally inconsistent. Write $\mathbf{y}_\beta = X\beta = \begin{pmatrix} \beta_0 + \beta_1 x_1 \\ \vdots \\ \beta_0 + \beta_1 x_4 \end{pmatrix}$ for the predicted response values under the given model using parameter vector β. Choosing different parameters β will in general yield different predicted response $\mathbf{y}_\beta$. When the matrix equation is inconsistent, there is no choice of parameter vector β to entirely eliminate the *error* $\epsilon = \mathbf{y} - \mathbf{y}_\beta$. Instead, we choose to minimize its magnitude, or equivalently its square-magnitude $\|\epsilon\|^2$. The expression $\|\epsilon\|^2$ is a sum of squares. What are the terms being added up? How you do interpret them in words?

We have

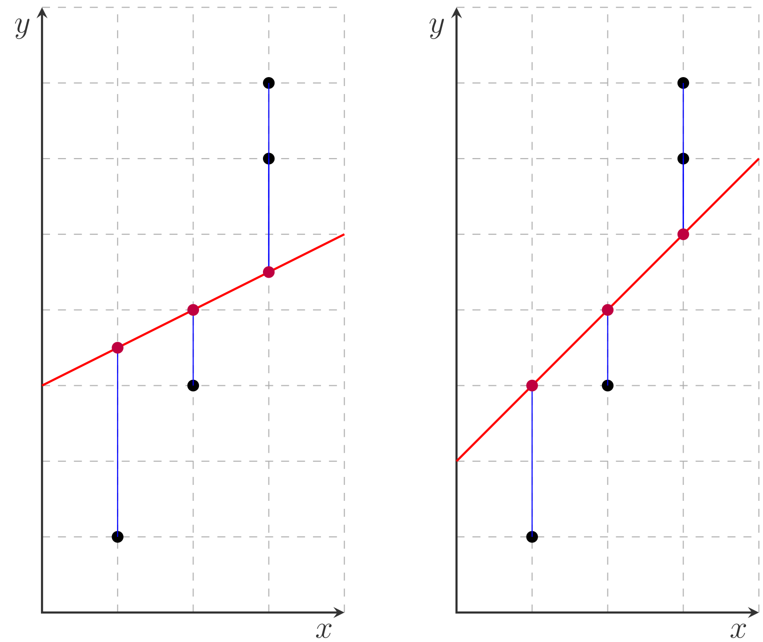
$$\|\epsilon\|^2 = \|\mathbf{y} - \mathbf{y}_\beta\|^2 = \sum_{i=1}^{4} (y_i - (\beta_0 + \beta_1 x_i))^2$$

a sum of square deviations of the predicted response (using parameter vector β) from the observed response for each sample.

(3) The plots below show the data (in black) and the line (red) generated by the predictor $\beta_0 + \beta_1 x$ for two different choices of β_0, β_1. (Left plot $(\beta_0, \beta_1) = (3, 1/2)$ and right plot

$(\beta_0, \beta_1) = (2, 1)$) For each of the plots, provide an interpretation for the terms in the sum of squares in $\|\epsilon\|^2$.

Graphically, the square deviations between predicted and observed response values are the square of the vertical lengths from the various data points to the line $y = \beta_0 + \beta_1 x$ determined by the choice of parameter vector β.



(4) In lecture we learned that $\|\mathbf{y} - \mathbf{y}_\beta\|^2 = \|\mathbf{y} - X\beta\|^2$ is minimized by least-squares solutions $\hat{\beta}$ to the matrix equation $X\beta = \mathbf{y}$. This explains the origin of the name “least-squares.” Thus, choosing parameter vector β will minimize the square-magnitude of the error between observed and predicted values of the response.

Our original model posited y as a linear combination of the *predictors* $1, x$. Suppose instead we model y as a linear combination of predictors $1, x, x^2$, i.e. $y = \beta_0 + \beta_1 x + \beta_2 x^2$ for some parameters $\beta_0, \beta_1, \beta_2 \in \mathbb{R}$. Write $\mathbf{p}_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, $\mathbf{p}_1 = \begin{pmatrix} x_1 \\ \vdots \\ x_4 \end{pmatrix}$, and $\mathbf{p}_2 = \begin{pmatrix} x_1^2 \\ \vdots \\ x_4^2 \end{pmatrix}$. If the data fit the model perfectly, what is the relation among the tuples $\mathbf{y}, \mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2$? Can you write this relation as a matrix equation involving the unknown tuple $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$?

If the model were a perfect fit, we could find parameters $\beta_0, \beta_1, \beta_2 \in \mathbb{R}$ such that

$$\mathbf{y} = \beta_0 \mathbf{p}_0 + \beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 = \begin{pmatrix} \mathbf{p}_0 & \mathbf{p}_1 & \mathbf{p}_2 \end{pmatrix} \beta = X\beta$$

where the columns of the model matrix X are the tuples $\mathbf{p}_i$. More explicitly,

$$\begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_4 & x_4^2 \end{pmatrix} \beta = \begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix}$$

(5) Again, generally the resulting system is inconsistent, and we must settle for a least-squares solution $\hat{\beta}$. Note that because the model changed, the column space of the model matrix changed, yielding new projection, error vector, and parameter vector. Our framework is sufficiently flexible to incorporate any well-defined predictor. For example, we could posit a model $y = \beta_0 + \beta_1 x + \beta_2 \sin x$ for parameters $\beta_i \in \mathbb{R}$. We can also change the response we are attempting to model, e.g. $\ln y = \beta_0 + \beta_1 x + \beta_2 \sin x$. What are the associated matrix equations for these two models?

The associated matrix equations are respectively

$$\begin{pmatrix} 1 & x_1 & \sin x_1 \\ \vdots & \vdots & \vdots \\ 1 & x_4 & \sin x_4 \end{pmatrix} \beta = \begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix} \qquad \begin{pmatrix} 1 & x_1 & \sin x_1 \\ \vdots & \vdots & \vdots \\ 1 & x_4 & \sin x_4 \end{pmatrix} \beta = \begin{pmatrix} \ln y_1 \\ \vdots \\ \ln y_4 \end{pmatrix}$$

The crux is that the response is modelled as a linear combination of predictors. Though this is a significant restriction on the types of models considered, linear regression remains highly versatile. In statistics, one elaborates on the above framework by providing a stochastic model for the error ϵ, which in turn yields confidence intervals for the parameter estimates.