Neural Networks for Signal Processing - I

Prof. Shayan Srinivasa Garani

Department of Electronic Systems Engineering

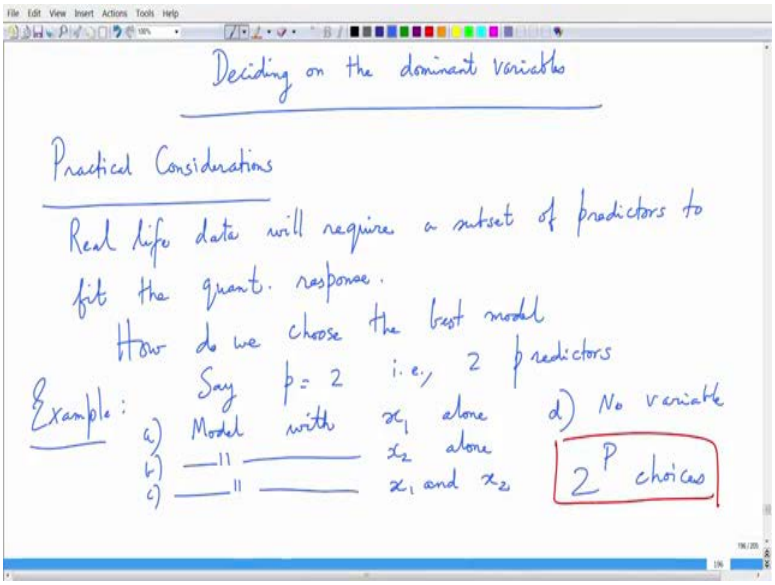
Indian Institute of Science, Bengaluru

Lecture – 14

Linear Regression 3

Let's delve deeper into linear regression. The first key topic we need to address is determining the number of dominant variables. In real-world data scenarios, not all variables hold statistical significance.

(Refer Slide Time: 00:52)



For example, let's consider a case where P = 2, meaning we have two predictors. In this scenario, we need to determine the best model out of the possible choices. Specifically, there are 2^P or 2^2 = 4 possible combinations.

With P = 2, we have four potential models to evaluate:

1. A model using only variable x_1,

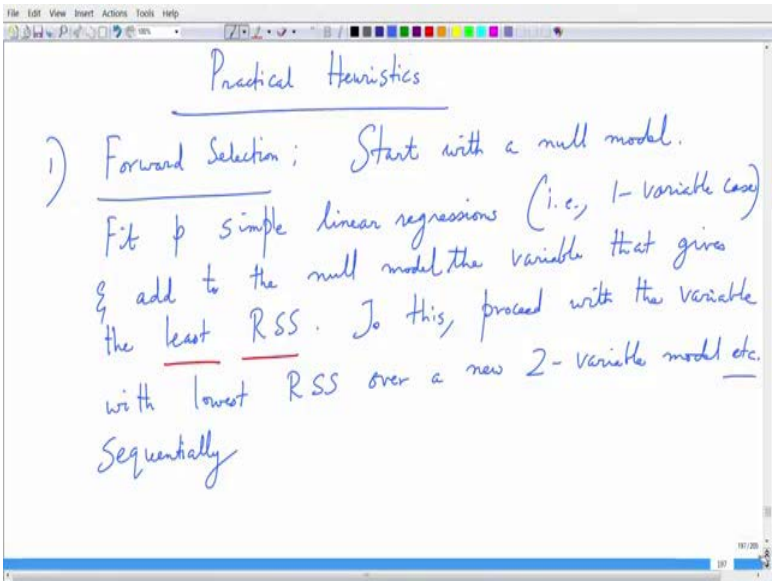
2. A model using only variable x_2,

3. A model using both x_1 and x_2,

4. A model with neither variable, which is usually not practical unless there is no meaningful relationship between the variables and the response.

In general, for model selection, we must consider 2^P possible models and identify the one that best fits the data. The choice of the optimal model can be guided by various information-theoretic criteria, such as the Deviance Information Criterion (DIC) or the Akaike Information Criterion (AIC). While we won't delve into all the specifics of these criteria here, it's crucial to recognize that model selection is a fundamental aspect of statistical analysis.

(Refer Slide Time: 02:15)



So, how do we select the best models in practice? One effective heuristic is the forward selection criterion. Here’s how it works:

We start with a null model, which means initially, we have no variables in the model. From this starting point, we fit p simple linear regressions, where each regression includes only one predictor variable. This means we evaluate each variable individually to see how well it explains the response.

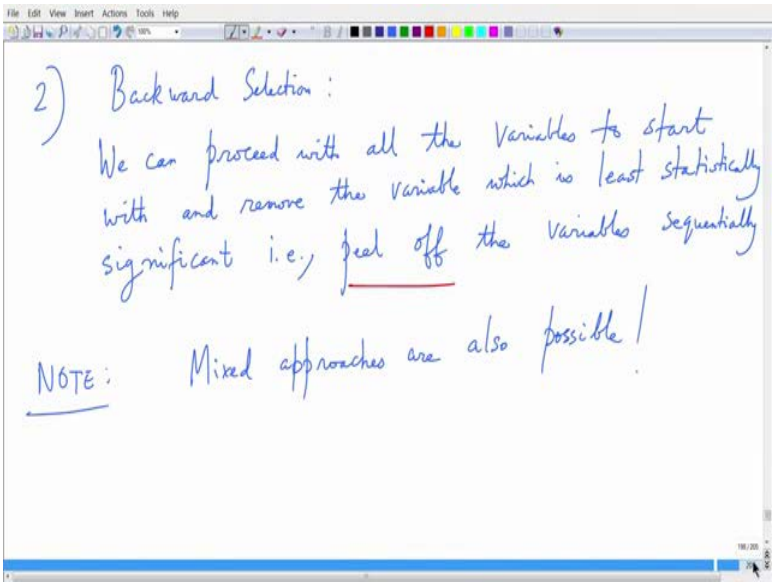
For each of these simple linear regressions, we calculate the residual sum of squares. We then select the variable that results in the smallest residual sum of squares and add this variable to our null model.

Next, we proceed by considering models with two variables. We add the new variable that,

when combined with the previously selected variable, results in the lowest residual sum of squares. In other words, we optimize over all possible two-variable models to find the combination that improves the model the most.

This process is repeated sequentially: at each step, we add a variable that contributes to the smallest residual sum of squares, refining the model incrementally. By following this approach, we continuously build a model that minimizes the residual sum of squares at every stage.

(Refer Slide Time: 03:57)

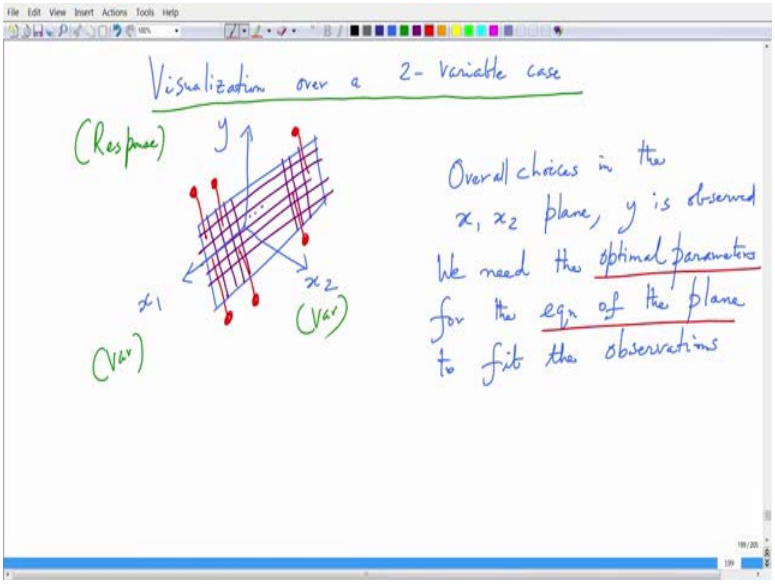


The backward selection process follows a conceptually similar approach to forward selection but in reverse. Instead of starting with a null model and adding variables one by one, we begin with a model that includes all the variables.

In backward selection, we initially include every variable in the model. Then, we identify and remove the variable that is the least statistically significant. To determine which variable to remove, we evaluate the error contribution of each variable. We then iteratively peel away the least significant variables until we arrive at a model that is optimized according to a specific criterion.

Additionally, mixed approaches are also feasible. For example, you can combine forward and backward selection processes, iterating between adding and removing variables to refine the model and achieve the best possible fit.

(Refer Slide Time: 05:06)



Having discussed the basic heuristics, it's crucial to pause and consider how we can visualize a 2-variable case. This visualization will aid in understanding the problem's geometry, particularly because visualizing problems with more than two variables is challenging. Thus, starting with simpler cases, like 1D and 2D, provides a clearer understanding of the geometric relationships.

Consider a scenario where we have two predictor variables, x_1 and x_2, and a quantitative response variable, y. In this context, y is our response variable, while x_1 and x_2 are the predictors. Visualize this as a plane defined by x_1 and x_2 where the response y is observed across various points on this plane.

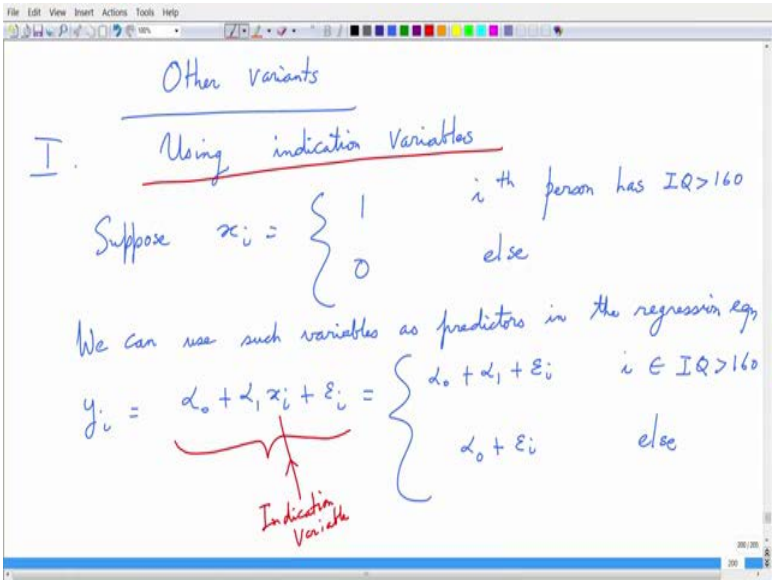
For continuous data, you'll see a range of y values for different combinations of x_1 and x_2. The task is to determine the optimal parameters for a plane that best fits these observations. In this visualization, the red sticks represent the response y corresponding to each pair of x_1 and x_2. I've discretized the x_1 - x_2 plane into a grid where each point represents a tuple (x_1, x_2) and its corresponding y.

Given these observations, the goal is to fit the optimal parameters for the plane equation that best matches the data. The matching criterion could be the residual sum of squares, an anomaly estimate, or any other appropriate criterion we choose to adopt.

Now, let's explore another variation of the linear regression problem involving indicator

variables.

(Refer Slide Time: 07:26)



For instance, suppose x_i is an indicator variable in this case, which takes the value of 1 if the i -th person's IQ is greater than 160, and 0 otherwise. Imagine we have a Mensa test to identify individuals with exceptionally high IQs. We might use this indicator to predict their future performance, perhaps whether they could achieve extraordinary accomplishments, like winning a Nobel Prize. In this scenario, y_i represents this prediction.

In the regression model, we can use such indicator variables as predictors. For example, if we model y_i as:

$$y_i = \alpha_0 + \alpha_1 x_i + \epsilon_i,$$

where x_i is the indicator variable. This model can be simplified based on the value of x_i. Specifically, if $x_i = 1$ (indicating an IQ greater than 160), the model simplifies to:

$$y_i = \alpha_0 + \alpha_1 + \epsilon_i.$$

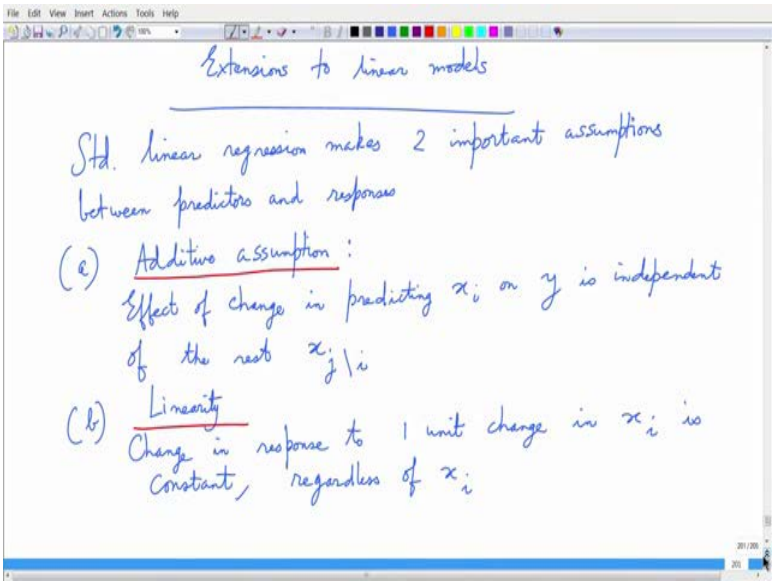
On the other hand, if $x_i = 0$ (indicating an IQ not greater than 160), the model simplifies to:

$$y_i = \alpha_0 + \epsilon_i.$$

This illustrates how indicator variables can be used to introduce different conditions or

levels into the model. You can include not just one, but multiple indicator variables to represent various levels or conditions, and set up the model accordingly.

(Refer Slide Time: 09:15)



There are several extensions to linear models worth considering. In the standard linear regression problem, two key assumptions are made about the relationship between predictors and responses:

1. Additivity Assumption: This assumption states that the effect of a change in predictor x_i on the response y is independent of the other predictors x_j (where $j \neq i$). In other words, the influence of x_i on y does not depend on the values of the other variables; it is simply additive.

2. Linearity Assumption: This assumption implies that the change in response due to a one-unit change in x_i is constant, regardless of the value of x_i. Essentially, the impact of x_i on y is linear and uniform across different levels of x_i.

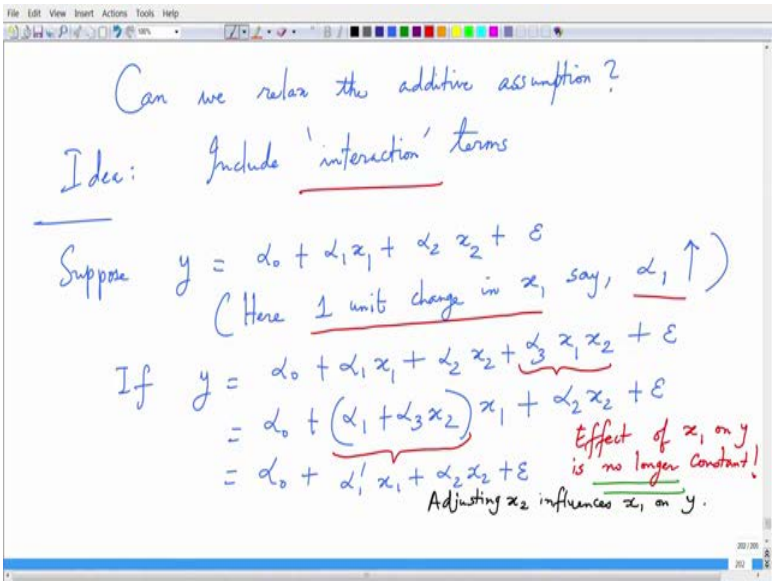
These assumptions can be more clearly understood when we express them in the equations used in linear regression.

Now, the question arises: Can we relax the additive assumption by incorporating interaction terms? To understand this better, let's explore it through an equation.

Consider the equation for a quantitative response y :

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \epsilon$$

(Refer Slide Time: 10:22)



In this model, a one-unit change in x_1 results in an increase in y by α_1. This indicates that the effect of x_1 on y is solely determined by α_1, independent of x_2.

To relax this assumption and account for potential interactions between predictors, we can include interaction terms in the model. The updated equation becomes:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 (x_1 x_2) + \epsilon$$

Here, α_3 represents the interaction term between x_1 and x_2.

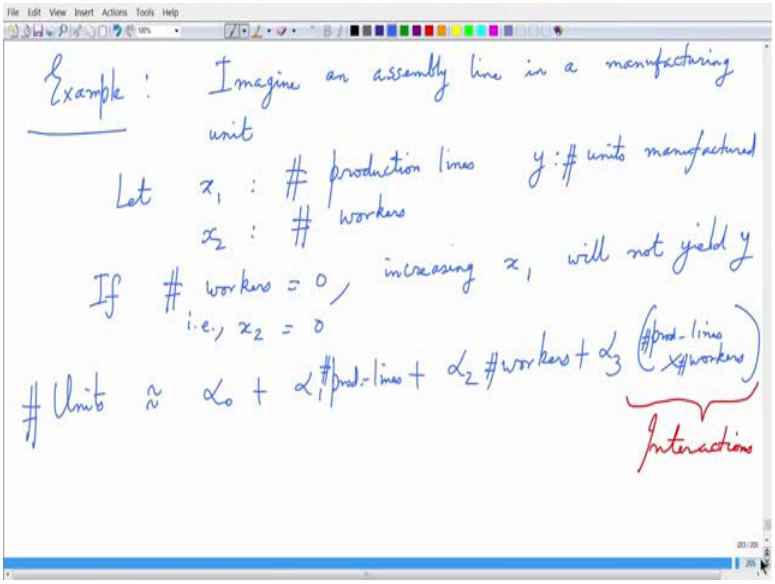
Let's simplify this expression:

$$y = \alpha_0 + (\alpha_1 + \alpha_3 x_2) x_1 + \alpha_2 x_2 + \epsilon$$

In this formulation, the coefficient of x_1 is now $\alpha_1 + \alpha_3 x_2$. This means that the impact of a one-unit change in x_1 on y is not constant; it depends on the value of x_2. The interaction term $\alpha_3 (x_1 x_2)$ indicates that the effect of x_1 on y is influenced by x_2.

Therefore, unlike in the previous model where the effect of x_1 was constant, in this new model, the change in y due to x_1 depends on the value of x_2. This highlights how the interaction between x_1 and x_2 affects the response y.

(Refer Slide Time: 13:26)



Let's explore this concept through a concrete example. Consider an assembly line in a manufacturing unit where we have two variables: x_1 and x_2. Here, x_1 represents the number of production lines, x_2 represents the number of workers, and y denotes the number of units manufactured.

It is intuitive to understand that if there are no workers (i.e. $x_2 = 0$), increasing the number of production lines will not result in more units being manufactured. Without workers, or in this case, even if we consider automated systems as workers, the production lines alone cannot generate output. This reveals an interaction effect between x_1 and x_2.

To model this scenario, we can express the number of units manufactured y as:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 (x_1 \times x_2) + \epsilon$$

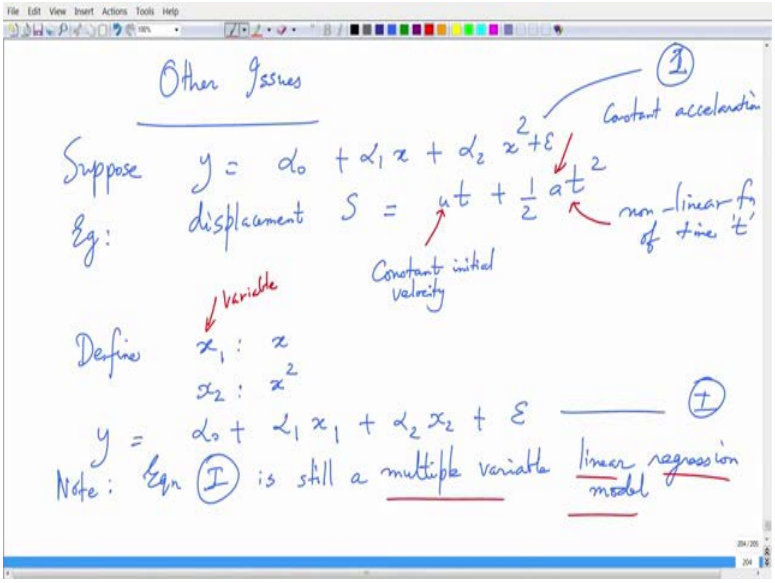
In this equation:

- α_0 is the intercept,
- α_1 is the coefficient for the number of production lines,
- α_2 is the coefficient for the number of workers,
- α_3 represents the interaction effect between the number of production lines and the number of workers.

This example illustrates how interaction terms can be used to capture the combined effects

of multiple variables. Besides interactions, other variations include indicator functions and non-linear interactions. Moreover, the scope of linear regression models extends beyond these examples, including considerations of non-linearity with respect to individual variables.

(Refer Slide Time: 15:15)



A classic example of non-linearity in equations is found in basic physics: the displacement s is given by the formula $s = ut + \frac{1}{2}at^2$. Here, u represents the constant initial velocity, a is the constant acceleration, and t is time. The displacement s is a quadratic function of time, featuring both a t term and a t^2 term.

In the context of regression, where our goal is prediction, it's worth asking if we can apply linear regression techniques to such quadratic relationships. To work with this, we need to express the quadratic function in a form compatible with linear regression.

Let's generalize this idea. Consider the equation:

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2$$

Here, let's denote x_1 as x and x_2 as x^2. Thus, we can rewrite our equation as:

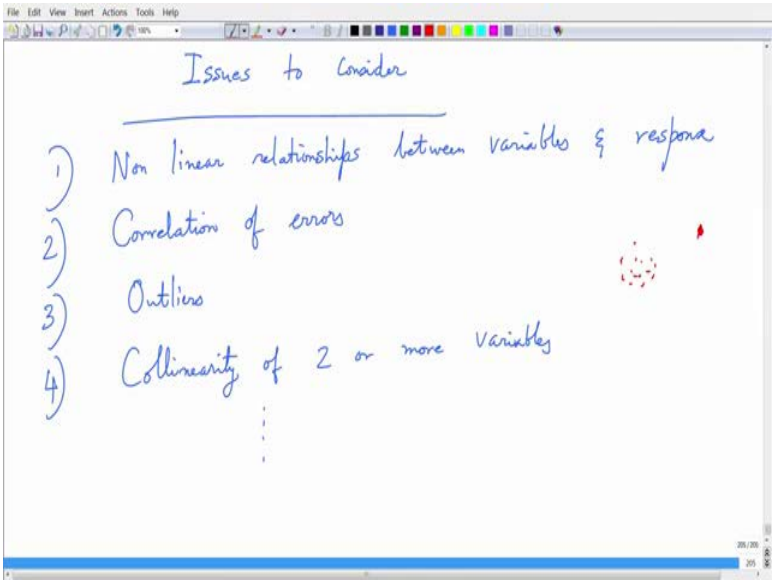
$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \epsilon$$

This setup is now akin to a multiple linear regression model, where x_1 and x_2 are our

predictor variables, and the original quadratic problem has been transformed into a two-variable linear regression problem.

By applying such transformations, we can leverage the standard linear regression techniques and tools to address problems that involve quadratic or other non-linear relationships. This approach allows us to use linear regression methods to model and predict even when the underlying relationship is non-linear.

(Refer Slide Time: 17:16)



In practical applications, there are several key considerations when working with linear regression models, which are worth discussing qualitatively. While an entire semester could be dedicated to these issues, here are the main points to keep in mind:

1. Modeling Non-Linear Relationships: Linear regression assumes linear relationships between predictors and the response. However, many real-world scenarios involve non-linear relationships. It is crucial to account for these non-linearities in your model.

2. Correlation of Errors: Errors in predictions can be correlated, which can affect the accuracy of the model. Handling these correlations appropriately is essential for improving model performance.

3. Outliers: Outliers are data points that deviate significantly from the norm. For example, if most data points are clustered closely together, but one point is far from this cluster, it may be an outlier. Deciding how to handle these outliers, whether to ignore them or assign

them specific weights, can significantly impact model results.

4. Multicollinearity: This refers to the situation where two or more predictor variables are highly correlated with each other. For instance, if pH is related to ambient temperature, or soil moisture is influenced by temperature, this collinearity can complicate the regression analysis. Addressing multicollinearity is crucial for making accurate predictions.

These practical considerations are vital for ensuring the robustness and accuracy of linear regression models. To delve deeper into these topics and enhance your understanding, I encourage interested students to explore textbooks on statistics or statistical modeling.

With these points covered, we'll conclude this topic and move on to logistic regression and explore aspects of maximum likelihood in the context of linear regression in the later parts of our study.