

Consider a dataset $D = (x_i, y_i)^N$ of N data points, where $x_i = (x_{i1}, x_{i2}, \dots, x_{iM})$ is a feature vector with M features, and y_i is the target, i.e., the response, variable. Let x_j denote the j th variable in feature space. A typical linear regression model can then be expressed mathematically as:

$$\begin{array}{ccccccc}
 & & \text{slope} & & & & \\
 & & \text{term for} & \text{input for} & & & \\
 \text{target} & & \text{feature_1} & \text{feature_1} & & & \\
 \backslash & & \underbrace{\hspace{1.5cm}} & \underbrace{\hspace{1.5cm}} & & & \\
 y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_M x_M
 \end{array}$$

This model assumes that the relationships between the target variable y_i and features x_j are linear and can be captured in slope terms $\beta_1, \beta_2, \dots, \beta_M$.