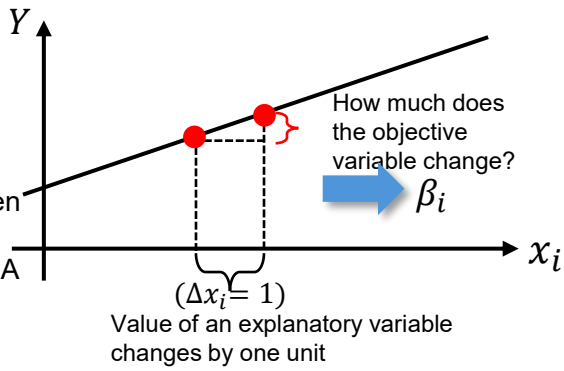
Contribution Analysis with Multi-Sigma®

1. Analysis of Variable Relationships Using Traditional Statistical Methods

In recent data analysis using machine learning, there has been growing attention on achieving high predictive accuracy. However, explicitly illustrating the underlying mechanism of behind the observed phenomena remains a challenging task. In contrast, traditional statistical analysis typically models the direct influence of explanatory variables on the objective variable. Among such methods, linear regression is a representative technique that clearly reveals the relationship between each explanatory variable and the objective variable. In the most fundamental approach, multiple linear regression, the relationship between multiple explanatory variables and a single objective variable is modeled. The regression coefficients serve as quantitative indicators of how much the objective variable is expected to change when a given explanatory variable increases by one unit, assuming all other explanatory variables are held constant.

$$Y = \beta_0 + \beta_1 x_1 + \boxed{\beta_2 x_2} + \beta_3 x_3 + \cdots + \varepsilon$$

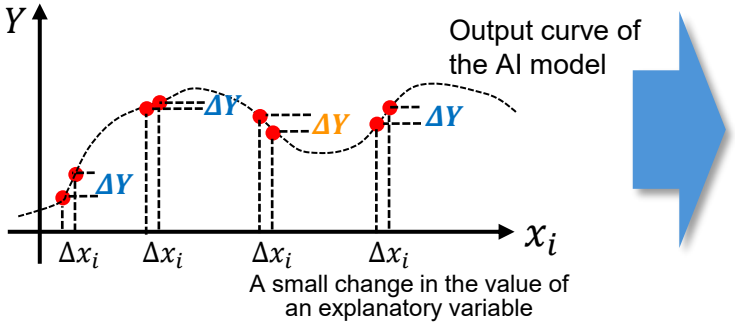
When x_2 changes by one unit, Y changes by β_2

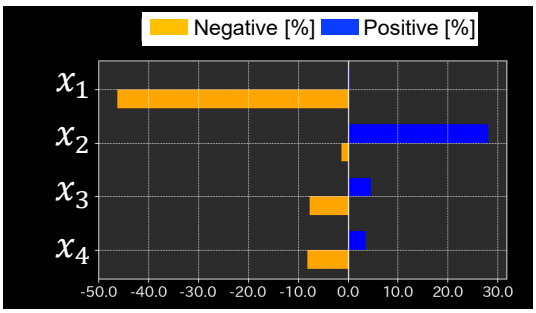


Then, how can one systematically capture the relationships between inputs and outcomes in data analysis using machine learning models such as neural networks or Gaussian process regression? A solution to this challenge is provided by the contribution analysis functionality of Multi-Sigma®.

2. Contribution Analysis using Multi-Sigma®

Sensitivity analysis is an effective approach for understanding the relationship between explanatory variables and the objective variable in machine learning models. This method involves observing how the predicted value of the objective variable changes when the value of a specific explanatory variable is slightly altered.
In Multi-Sigma®, this sensitivity analysis serves as the foundation for a process in which small, incremental changes are applied sequentially to multiple explanatory variables to evaluate their cumulative impact on the objective variable. Through this process, even machine learning models that appear to function as black boxes can be analyzed to intuitively and quantitatively reveal the extent to which each variable contributes to the outcome.





A key advantage of this approach is that it allows the evaluation of underlying relationships within the data without relying on the internal structure of machine learning models. Regardless of model type, contribution analysis can be performed. Even when positive and negative effects coexist, their cumulative impact can be quantitatively assessed by aggregating them sequentially. Multi-Sigma® also offers visualization features that help users intuitively grasp the results. This approach is broadly applicable to areas such as data-driven business decisions and the exploration of mechanisms in basic research.