ABSTRACT

In a data analysis system for determining a correlation model between biological conditions and expression levels of a plurality of genes and/or quantities of intracellular substances, A data set is constructed by using the biological conditions or changes of the biological conditions probabilistically generated with time as object variables, and the expression levels of the plural genes and/or the quantities of intracellular substances as explanatory variables. The explanatory variables contained in the data are selected, and cross validation is calculated for a correlation model containing the selected explanatory variables and object variables to assess the results. Selection of the explanatory variables, calculation of cross validation and discriminative assessment of the result is repeated until no further improvement of the cross validation is observed to determine a partial least square model. This permits an effective processing method to provide multivariable gene information.