In ancient times, scientists conducted their research blindly. The work plan was based solely on their own understanding of the subject and did not rely on quantitative characteristics. But, everything changed with the advent of Chemometrics. Chemometrics is a scientific discipline located at the intersection of chemistry and mathematics, the subject of which is mathematical methods for studying chemical phenomena. It includes such areas of mathematics and programming as machine learning, data science, big data, and deep learning.

Chemometrics helps solve regression and classification problems. There are many types of regression. However, before you start training, you need to prepare the data. For example, if the data contains non-numeric data, then you need to interpolate the sets with a smooth piecewise cubic function. This function is given by the Hermite interpolation polynomial. There may also be undesirable peaks in the experimental data that also need to be interpolated. The use of the Hermite polynomial for interpolation is not accidental. This polynomial contains not only the polynomial of the desired degree, but also terms that take into account the value of the derivative at the nodes. Unfortunately, the Lagrange interpolation polynomial does not give us such advantages. The advantage of the Hermite polynomial is that it gives a strictly monotone function on the interval between two nodes. For comparison, spline interpolation does not always produce a monotone function between two nodes. Then the algorithm differs significantly depending on the goals being pursued. Using the example of a numerical object regression problem, I will explain the General strategy. After preliminary data processing, data analysis follows. First, cross validation is performed. This is a procedure for dividing a data set into test and training sets. I will consider the learning situation with the test values of the desired component. In this case, the data consists of four matrices. Two feature matrices and two value matrices for the desired components. However, it happens that not all signs affect the change of the desired value. Some features are unnecessary, and the principal component method is used to determine these extra features and noise in the data. Its essence is to move to a new space, where the optimal dimension of the new space is selected by projecting and calculating the error. Using the principal component method, for example, it is possible to reduce the number of features by several hundred times. Then the model is trained on training datasets. After training, the model is tested on test datasets, because the problem of retraining is common. This is when the model perfectly predicts training sets, but begins to lie on test sets.

I think next year, new ways of applying Kruskal tensors and Kruskal regression in chemometry will be discovered.