# PCA transform

We use the PCA on the “pcr results” measurements. Since there are 16 such measurements, we want to find the most dominant Principal Components.

First we full the missing data with the median of each “pcr result” caregory, since they have gaussian-like distribution.

Then, we use the Scaler function to normalize the values of all the measurements.

Then, we define the number of Principal Components we want to save. We choose 5 arbitrarily, later this number will be adjusted according to the classifier performance.

We project the feature vectors of “PCR results” onto the 5 Principal Components. We get 5 new features, which are in fact linear transformation of the original features onto the 5 Principal Components.

So we do want to **save all 16 pcr results features,** but those will be transformed into only 5 new features, which are the values on the 5 biggest Principal Components.