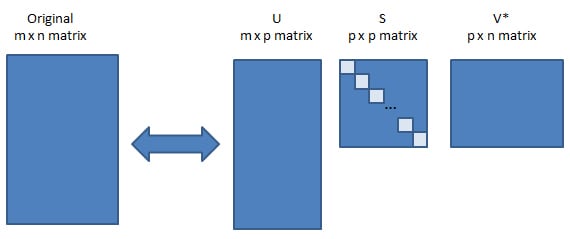
SVD, or Singular Value Decomposition, is one of several techniques that can be used to reduce the dimensionality, i.e., the number of columns, of a data set. Why would we want to reduce the number of dimensions? In predictive analytics, more columns normally means more time required to build models and score data. If some columns have no predictive value, this means wasted time, or worse, those columns contribute noise to the model and reduce model quality or predictive accuracy.

SVD is an algorithm that factors an *m x n* matrix, *M*, of real or complex values into three component matrices, where the factorization has the form *USV\**.

*U* is an *m x p* matrix. *S* is a *p x p* diagonal matrix. *V* is an *n x p* matrix, with *V\** being the transpose of *V*, a *p x n* matrix, or the conjugate transpose if *M* contains complex values.

The value *p* is called the rank. The diagonal entries of *S* are referred to as the singular values of *M*. The columns of *U* are typically called the left-singular vectors of *M*, and the columns of *V* are called the right-singular vectors of *M*.

Consider the following visual representation of these matrices:



One of the features of SVD is that given the decomposition of *M* into *U*, *S*, and *V*, one can reconstruct the original matrix *M*, or an approximation of it.

The singular values in the diagonal matrix *S* can be used to understand the amount of variance explained by each of the singular vectors.