For certain kinds of objects, it is not clear how they should be best represented as fixed-size feature vectors, such as those with variable lengths, sizes, and shapes, as well as complex three-dimensional geometry. One approach in order to figure out how to represent these objects is to define a generative model for the data and then use the inferred latent representation or the parameters of the model as features and use these features in standard methods. Another approach is to assume that there is some way of measuring the similarity between objects which doesn't require preprocessing them into vector format, using some measure of similarity between objects which we call a kernel function.

A kernel function is defined as a real-valued function of two arguments, $\kappa(x, x') \in R$, for $x, x' \in X$. This function is typically symmetric and non-negative so it can be interpreted as a measure of similarity, but these characteristics are not required.

When performing document classification or retrieval, it's useful to have a way to compare two documents, given as x_i and $x_{i'}$. This can be done using the bag of words representation and the cosine similarity, which gives us a quantity measuring the cosine of the angle between x_i and $x_{i'}$ with a value between 0 and 1, where 0 would mean that the vectors are orthogonal and therefore have nothing in common. However, this method does not work very well. The reasons why are because if the two documents have any word in common they will be deemed similar, even if it is some popular word that they have in common, such as stop words like "the" and "and". Secondly, if a discriminative word appears many times in one of the documents, the similarity will become artificially boosted, even though once a word is used in a document it is very likely to be used again. The performance of the overall approach can be improved using preprocessing, which can be used to replace the word count vector with a new feature vector called the term frequency inverse document frequency, abbreviated as the TF-IDF.

Some methods will require that the kernel function satisfy the requirement that the Gram matrix must be positive definite for any set of inputs. Such kernels are called Mercer kernels, or also positive definite kernels. There are also Matern kernels, which are commonly used in Gaussian process regression.