Unsupervised Learning

As opposed to Supervised Learning, in unsupervised learning, the dataset is unlabeled and hence there is no right or wrong answer. The objective of unsupervised learning is thus to extract knowledge from the data, for instance, by discovering hidden patterns. In this video, we'll discuss two kinds of unsupervised learning, clustering and data transformation. Clustering concerns partitioning data into groups called clusters. The goal is to split up the data in such a way that points within a single cluster are very similar, and points in different clusters are different. For instance, by clustering an image dataset, we may end up with four clusters. Images of people, of animals of plants, and of buildings. But we could also end up with three completely different clusters, photos that were taken on a sunny day, on a cloudy day, or on a snowy day. K-means clustering is one of the simplest and most widely used clustering algorithms. The main idea is to find cluster centers that are representative of certain regions of the data. This is done by alternating between two steps, assigning each data point to the closest cluster center, and then setting each cluster center as the mean of the data points that are assigned to it. Let's illustrate how the algorithm works on a fictitious example. Suppose our data looks like this, and we're looking for three clusters. We initialize the algorithm by randomly setting three data points to cluster centers, like this. Next, the assignment phase takes place, where each data point is assigned to the closest cluster center. After the assignment phase, we recompute the cluster centers. We go through another couple of rounds of point assignment and center recalculation, until there are no further changes in point assignment. These are the three clusters that we have discovered. It is worth mentioning that the number of clusters, k, is an input parameter, and if we don't set it to an appropriate value, we may get poor results as demonstrated here. Clustering is widely used in medicine and health care, for instance, to identify similar patients based on their attributes and explore potential treatments. Apart from clustering, we may carry out unsupervised learning so as to transform a dataset, for instance, to compress the data and find a representation that is more informative for further data processing. This is commonly referred to as dimensionality reduction and feature extraction. A simple method that is commonly used for dimensionality reduction is principal component analysis, or PCA, as it's typically called. The goal here is to reduce the number of features, while retaining as much useful information as possible. Let's explain how PCA works for a dataset with two features. In other words, for two-dimensional data. Suppose that our dataset looks like this. The first thing that the algorithm does is find the direction of maximum variance. This is the direction of the data that contains most of the information, or in other words, the direction along which the features and most correlated with each other. We call this component one. Next, the algorithm finds a direction orthogonal to component one that has a maximum variance. This is component two. Components one and two are the principal components of our dataset, which are the main directions of variance in the dataset. We can see that if we

return these dimensions, we're not going to lose any information. It's simply a rotation of our original dimensions our original axes. So essentially, we've transformed or reconstructed our features. After identifying the principle components, we can use PCA to carry out dimensionality reduction by dropping one, or more components that have the smallest maximum variance. In our two-dimensional dataset this means retaining only component one. This is the reduction that gives the minimum error in the case that we decide to reconstruct the original data set. In other words, PCA allowed us to reduce the number of dimensions in our original data set from two to one, and get a transformed dataset that retains as much useful information as possible. PCA is commonly used to speed up supervised learning. So, if we have a training set with a large number of features, we can use PCA to reduce the number of features in each example of the training set. PCA has been widely used in the medical field. For instance, to compress medical imaging data. As we have discussed in this video, unsupervised learning allows us to extract knowledge from unlabeled data. A big challenge here is evaluating whether the algorithm learn something useful or not. So, performance here is often subjective and domain-specific. Therefore, unsupervised learning is often carried out in an explorative setting or as a pre-processing step for supervised learning