Unsupervised and Supervised Learning Problems

Let's look at some Data analysis problems and state whether they are supervised or unsupervised problems:

Problem 1:

Suppose we have measured the concentration in the air of a certain type of pollution. We have measured the amount of pollution at different locations, under different weather conditions at different times of the day and for different days of the week.

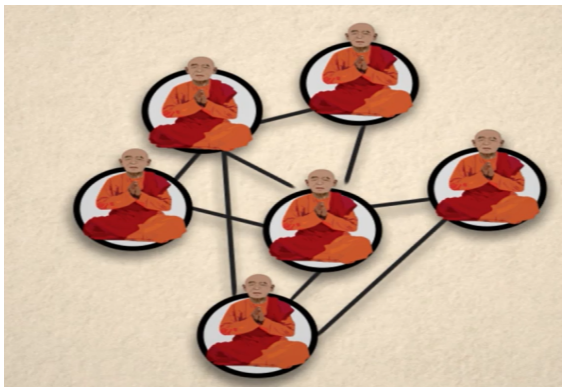
Now we would like to predict the air pollution concentration at a new location where weather and date are given.

This is an example of **Supervised Learning**. Here the data points are the collection of information about the surroundings like Location, Weather, Date & time. And the labels are air pollution concentration which is a numerical value.

Problem 2:

When Frank Simpson was a Ph.D. student in the 60's he spent some time in a Monastery. He followed the novices' training there and recorded their interactions with each other.

Now we want to discover what the friendships or social groups among the novice monks are.



This is an **Unsupervised Learning problem**. As here, we don't have any labels that tell us who belongs to which group and are there any social groups that exist.
Here we have to find the pattern in the data in an unsupervised way.

We have some similar kinds of problems like this one. For example: Detecting communities or groups in large social networks like Facebook or Twitter.

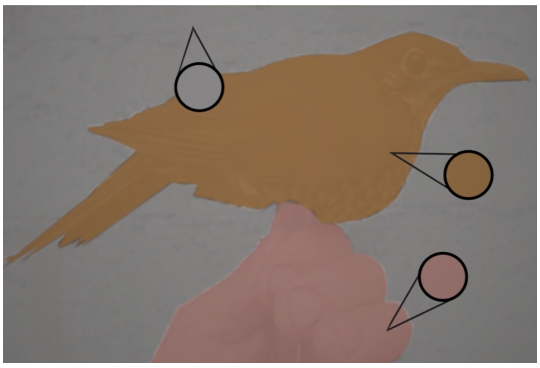
Problem 3:
NASA uses rocket boosters to launch spacecraft into orbit. It's expensive to build a whole new rocket booster for every slight tweak and it's difficult to recreate real atmospheric conditions in a wind tunnel for testing these out. So, NASA uses computer simulations, but it takes a long time to run the computer simulation for different rocket boosters specifications.

So, NASA would like to measure the simulated lift, for different rocket booster specifications and predict the simulated lift under other given specifications.

This problem is a **Supervised Learning Problem**. Here the data points are the rocket booster's specifications and the labels are the simulated lift.

Problem 4:

This problem is a little different than the previous problems.



Suppose, we have small cameras that we have installed to monitor bird's nests in New England. But, we don't have a lot of Bandwidth to send the video. So, we have to compress the recorded video before sending it.

We can think of this as an Unsupervised Learning Problem. If we consider the pixels at each timestamp as a data point, then we can group together the pixels which are similar in color and similar in time to achieve the compression.