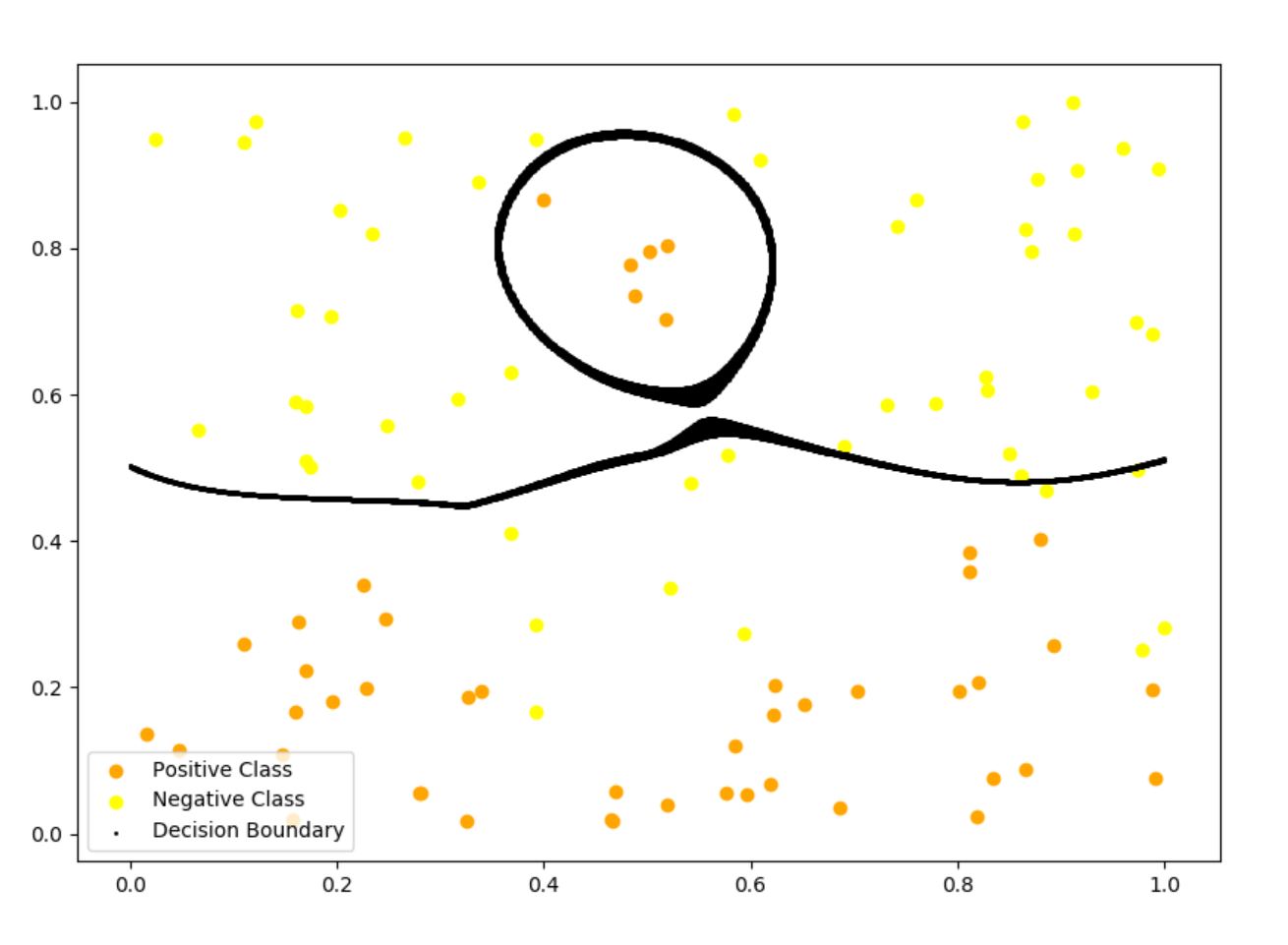
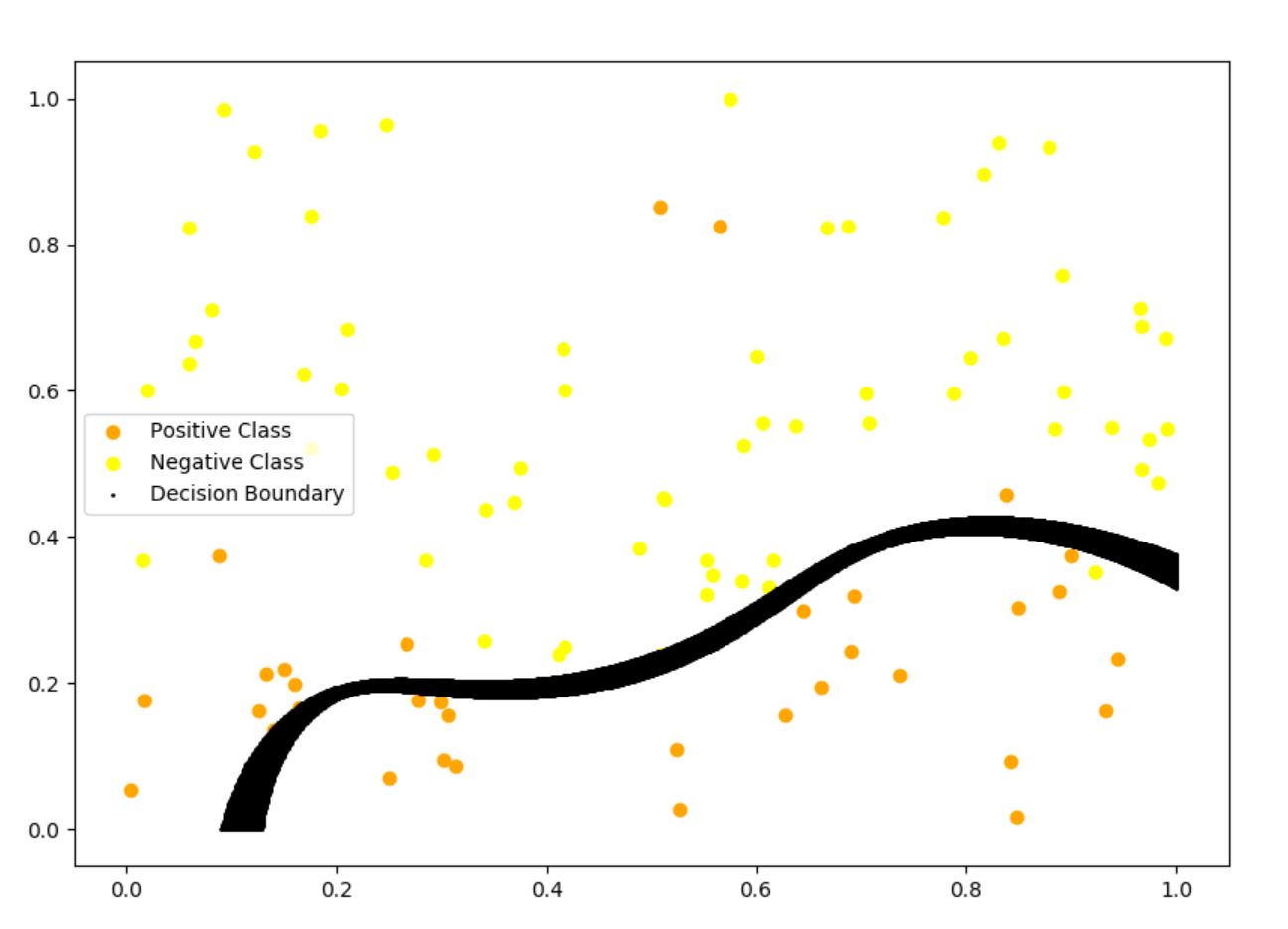


**With number of centers 20**

 **With number of centers 4**



**The correct choice of *k* is often ambiguous, with interpretations depending on the shape and scale of the distribution of points in a data set and the desired clustering resolution of the user. In addition, increasing *k* without penalty will always reduce the amount of error in the resulting clustering, to the extreme case of zero error if each data point is considered its own cluster. Source – Wikipedia.**

**Clearly, from the above plots, with more number of centers, we are able to separate the points more accurately. With just 4 centers, the margin or the decision boundary is able to separate the lower part of the points in the plane but not able to separate the above circle of points. With 20 centers, each class is getting 10 centers which is sufficient to accumulate most of the points.**

**Also, since the maximum number of negative class points (sun and mountain) would be present in the lower part of the plane, probabilistically, it is more likely for clusters to get accumulated in the lower part resulting in more cluster heads in the lower part of the plane. Thus, with lesser number of cluster heads, we can not get an accurate decision boundary for the points in the upper part (sun part). In our case, k = 20 is better than k = 4.**