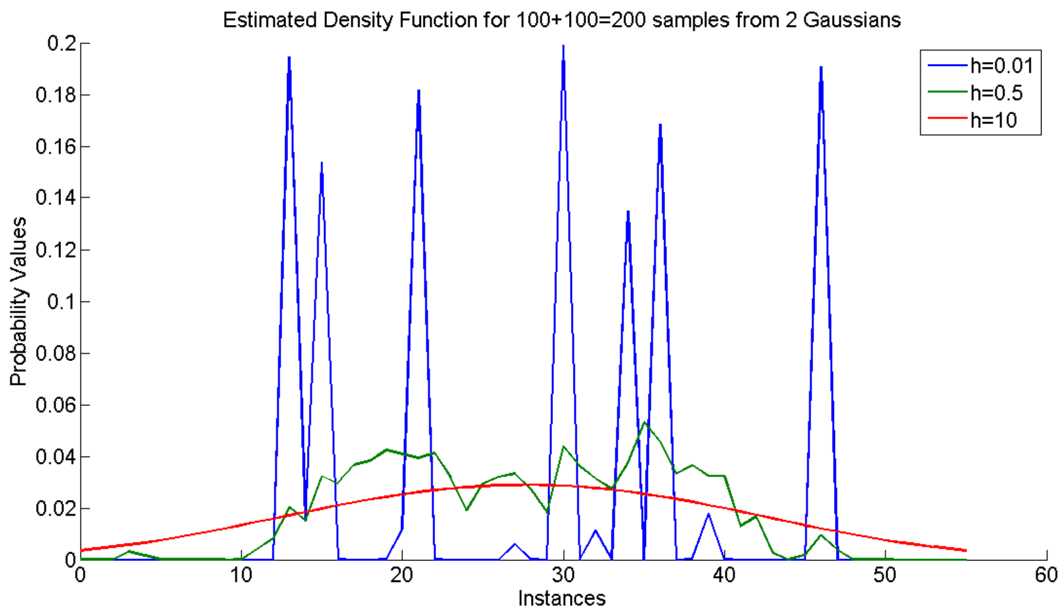
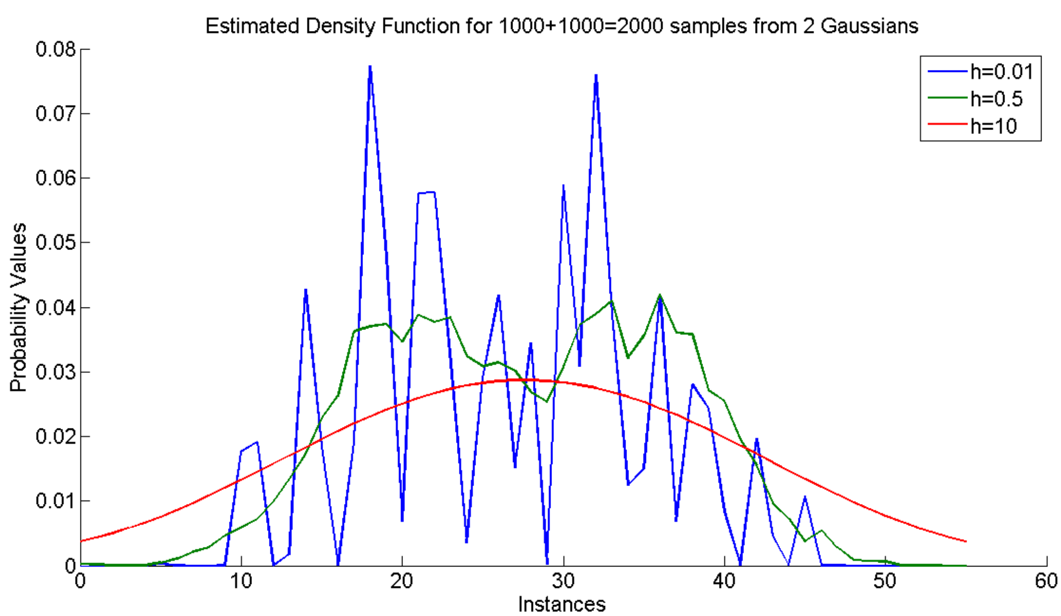
Question 1:

a) The estimated density functions for bandwidths of 0.01, 0.5 and 10 given 200 sample points (100 from each one of the normal distributions) are given in the figure below.



b) The estimated density functions for bandwidths of 0.01, 0.5 and 10, this time given 2000 sample points (1000 from each one of the normal distributions) are given in the figure below.



c)

i) Bandwidth effect: The bandwidth is the most important parameter in kernel density estimation. Depending on the size of bandwidth, we may have 3 conditions:

1) Over-smoothing
2) Under-smoothing
3) Good-fit

In both figures we have tried out different bandwidth values of 0.01, 0.5 and 10.

As we can see, h=0.01 in both figures (blue line) is a very small bandwidth hence resulting in under-smoothing. As we see, for h=0.01, most of the time no point is captured and probability value is zero and when we observe some points in the bandwidth we see sparks in the blue line. Therefore, h=0.01 is a very small bandwidth and cannot adequately model the points coming from two Gaussian distributions.

Another bandwidth we tried is h=0.5 (green-line). Among the bandwidths we tried, the green-line provides the good-fit condition. As we can see from both figures, with h=0.5 we are able to observe the two Gaussians that generated our sample points.

The last bandwidth we tried is h=10 (red line). H=10 is a too big bandwidth value for our sample points and it gives us the condition of over-smoothing. As we see in two figures, red line models the sample points as if they were coming from a single Gaussian distribution (over-smoothing), whereas in fact we had 2 Gaussians.

ii) Sample size effect: The sample size is another important factor in kernel density estimation. As we increase the number of sample points in the space it is likely that we will capture more points in our bandwidths and will better model the underlying density function. For example if we take a look at the blue line in both figures (h=0.01 or under-smoothing condition) we can see that the zero probability points have become less when we increased the sample size. That is due to the fact that, now we have more instances and more points fall into our small bandwidth value of 0.01.

If we take a look at the red line, we see that the red line in the first figure is a lot smoother (close to a line) when compared to the second figure. Since we have more instances in the second figure, the red line looks more like a Gaussian distribution. So we again observe the effect of more points in generating a better estimate.

Finally, if we consider the green line (good-fit condition) in two figures, we see that green line lets us observe the two Gaussians. However, the green line in the second figure is smoother in comparison to first figure and looks more like two Gaussians. That is due to the fact that more points mean more samples in the bandwidth and therefore more points help us estimate the underlying function better.