MACHINE LEARNING

Question 1 to 11.

Question	Answers
1	A
2	D
3	A
4	A
5	B
6	B
7	A
8	D
9	A
10	D
11	D

Question 12.

Is K sensitive to outliers?

Answer:

The k-means algorithm updates the cluster centres by taking the average of all the data points that are closer to each cluster centre. When all the points are packed nicely together, the average makes sense. However, when you have outliers, this can affect the average calculation of the whole cluster. As a result, this will push your cluster centre closer to the outlier.

An example, is the average of the salaries of the following people:

$50k, $20k, $35k, $65k and $1 Million

The average ends up being ($50k + $20k + $35k + $65k + $1MM) / 5 = $1170k / 5 = $234k.

If we did not have the $1MM outlier, the average would have been ($50k + $20k + $35k + $65k) / 4 = $170k / 4 = $42.5k.

Note that the two average results are wildly different from one another.

Given that k-means clustering is an unsupervised algorithm, it is up to the interpreter to determine whether this makes sense or not for a given data set. There are other clustering algorithms out there that are less sensitive to outliers. Depending on your application it may be worth using a different approach than the k-means algorithm.