Classification Module Assignment 1 Solution

1. The correct answer for the 1st question is option 4 or option d.

Here is the explanation for the same

1. This is true. If you use the training data set as the test set, then with one nearest neighbour, if given a point x, the nearest neighbour will be the exact same point and thus the error will be 0. For 5-NN, 0 is a lower bound.
2. False. Consider a 1d example. You have x\_{{train}} = (-5, -4, -3, -2, -1, 3) and y\_{{train}} = (0, 0, 0, 0, 0, 1). Now consider a new point with x=2 and y=0. Then this will have test error 100% for 1-nn and 0% for 5-nn.
3. No. Consider two classes, one is in the shape of a moon and the other surrounds the moon. Then the decision boundary will have approximately the shape of a moon.
4. This is true. At test, KNN needs to make a full pass through the entire data set and sort points by distance. The time needed thus grows with the size of the data.
5. The answer for 2nd question is in the Ipython notebook.