Software Development IV – Advanced .NET 420-411-DW
Lab Exercise 1 – K-Nearest Neighbors

At the point : would you predict that this student passed or failed the exam?

Your instincts probably tell you that the poor student most likely failed - because you noticed *clusters*.

K-Nearest Neighbours algorithm

This algorithm basically checks which training data points are close to the new point , and predicts its classification based on the neighbours. indicates the number of nearest neighbours who get a vote. Let's say we say that for the dataset above. Which are the three closest neighbours to ?

The easiest distance measure is *Euclidean* distance. In our example, we have two variables, or dimensions, or *features*. Recall, in two dimensions, the distance between point and point is:

4. Count the frequencies of the k labels. The highest frequency is the predicted label

Lab exercise

We will be looking at the Iris flower dataset and using it to classify a new flower. This is one of the first datasets that is typically used in machine learning exercises.

The flower dataset contains measurements of 3 different but related species of irises. For each species, there are 50 flowers which were measured; and each f