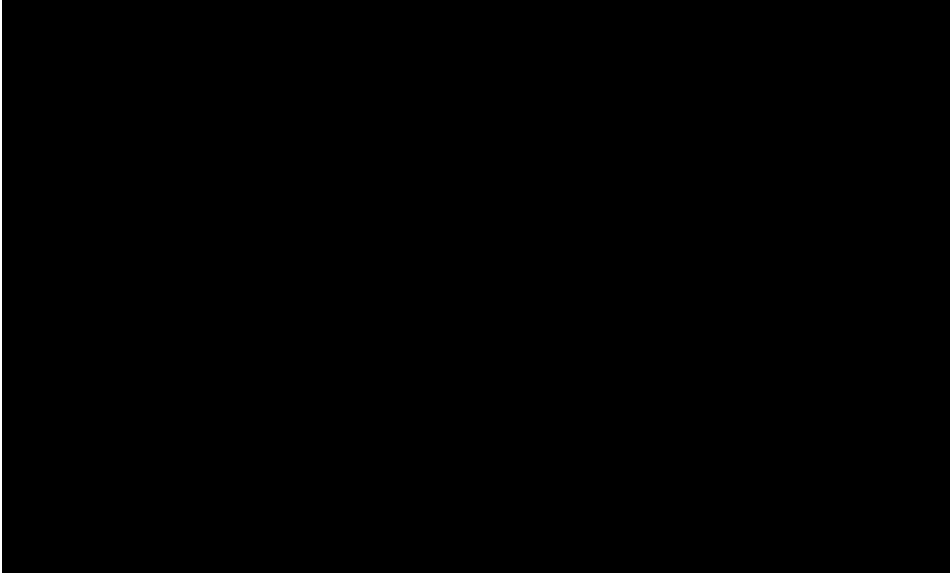


**Software Development IV    Advanced .NET 420-411-DW**  
**Lab Exercise 1    K-Nearest Neighbors**



At the point  $x$ : 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  $x$ , and predicts its classification based on the neighbours.  $k$  indicates the number of nearest neighbours who get a vote. Let's say we say that  $k=3$  for the dataset above. Which are the three closest neighbours to  $x$ ?

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  $(x_1, y_1)$  and point  $(x_2, y_2)$  is:

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