**1**. C­1=Senior, C2=Junior  
P(C1) = 5/11  
P(C2) = 6/11  
  
P(department = system | C1) = 2/5  
P(department = system | C2) = 2/6

P(salary <= 40k | C1) = 1/5  
P(salary <= 40k | C2) = 1/2

P(age = 31…40 | C1) = 3/5  
P(age = 31…40 | C2) = 2/6  
  
P(department=system, age=31…40, salary<=40k | C1) = 2/5 \* 1/5 \* 3/5 = 6/125  
6/125 \* P(C1) = 6/125 \* 5/11 = **0.0218**  
  
P(department=system, age=31…40, salary<=40k | C2) = 2/6 \* 2/6 \* 1/2 = 1/18  
1/18 \* P(C2) = 1/18 \* 6/11 = **0.0303**

The tuple belongs to the class **Junior**.

**2**. We want probability of student in this school being an undergrad and smokes. Because these are independent, we can simply multiply P(undergrad in this school) and P(undergrad who smokes) so: 0.25 \* 0.13 = 0.0325.

**3**. Support Vector Machines have a higher classification accuracy in high-dimensional space because it becomes easier to separate the datapoints with a hyperplane. Depending on the kernel function, transforming the training data to a higher dimension increases the distance of the two groups of datapoints that correspond to each class. Because of this, all the SVM algorithms need to do is to find the W and b in the equation W\*X + b for the given training dataset.   
  
Once the values of W and b have been found, the SVM algorithm essentially becomes a special case of a rule-based classifier. If W\*X + b >= 1, then it is class 1, else if W\*X + b <= 1, then it is class 2. These if…then’s are part of the definition of a rule-based classifier, thus the last step in the SVM algorithm may be considered a special case of a rule-based classifier with two rules.

**4**.

**5**. For example 1, the hierarchical clustering using minimum distance would produce two nicely separated clusters. The minimum distance approach can handle non-elliptical shapes. Since there are not many outliers and noise in example 1, the minimum distance hierarchical clustering will not be affected and perform well.

For example 2, k-means would work best. K-means performs well when the shapes of the clusters are hyper-spherical, which is the case in example 2. The data does not seem to be noisy nor there seems to be many outliers so this will have no affect on the algorithm.