Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Answer1:
Maximum margin classifier allows the two classes to be separable with maximum margin from the nearest support vectors of both the classes to the hyperplane that divides these two classes. This classifier doesn't allow any misclassification in order to make the model more generalizable.

Soft Margin Classifier also follows the same principle of maximum margin classifier to maintain maximum distance, however it is not so strict in data points segregation. It allows certain misclassification to happen as the real world data is always noisy and in tern make the model more generalizable. It controls this with the help of Hyperparameter C. Larger the value of C more misclassification is allowed.(answer based on theory explained by professor not sklearn library).

Question 2

What does the slack variable Epsilon (ε) represent?

Answer 2:

Slack variable Epsilon (ε) is used to control the misclassification. It tells where an observation is located with respect to the hyperplane and margin.

If $\varepsilon = 0$, then SVM model acts as Maximum margin classifier not allowing any misclassification.
If $\varepsilon > 1$, then SVM allows misclassification.
If $0 < \varepsilon < 1$, then observations get very close to the hyperplane, they violate the margin.

Question 3

How do you measure the cost function in SVM? What does the value of C signify?

Answer 3:

Cost function in SVM:

$min_{w,b}\ 1/2||w||^2 + C\sum_n\ \varepsilon_n$

Such that,

yn*(W^T*Xn) >= 1- ε n

ε = epsilon

C = some variable (Hyperparameter that controls epsilon)

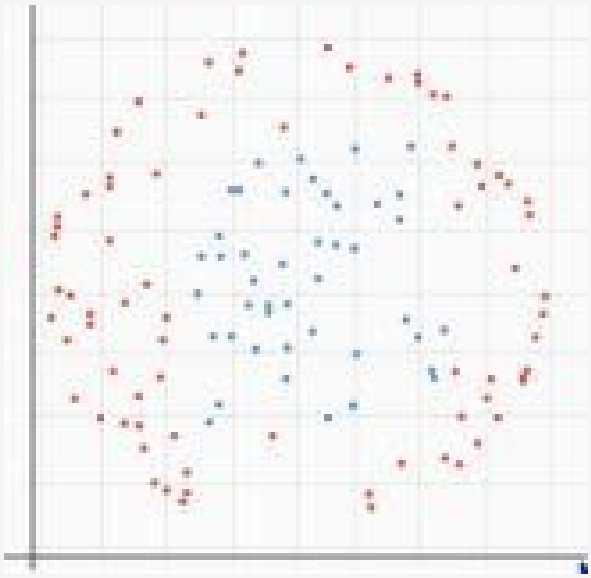
The C here controls the amount of misclassification the model can accommodate.

Higher the value of C, high is the cost of misclassification('Hard Margin') that leads to overfitting.

Lower the value of C, lower the cost of misclassification('Soft Margin') That leads to under fitting.

Note: This cost function aligns with the sklearn library and many other books, hence easy to remember.

Question 4



Given the above dataset where red and blue points represent the two classes, how will you use SVM to classify the data?

Answer 4:

As in the above figure, we can see that the observations are not linearly separable in this X,Y plane. We can either perform feature transformation if we know the appropriate function or we can make use Kernels and try out different variants like Linear, polynomial or RBF to see which one gives better accuracy.
So, in order to classify the observations we need to transform the features into a higher dimensional space.

Question 5

What do you mean by feature transformation?

Answer 5 : In the context of SVM, when a set of observations are not linearly separable through a line or a hyperplane. We transform the data into a different space which is usually high dimensional space using the kernel trick to make it possible to perform linear transformation. We can either find out $\phi(xi).\phi(xj)$ if we know the function ϕ and transform the vectors into a new feature space or use kernel function that calculates the dot product in a different space without even visiting it.