1. SVMs aim to maximize the \_\_\_\_\_ between classes.

The Margin. Or to be more specific, we want to maximize our hyperplane so that it is greater than some value rho (ρ):

2. Instead of representing samples as vectors, SVMs can represent pairs of samples as \_\_\_\_\_\_.

Slack Variables.

3. What is soft error? How does it relate to the cost parameter?

Soft errors occur when the Slack Variables are greater than 1.  When this occurs the variable is on the wrong side of the margin.

4. Name a few kernels available for use with support vector machines.

Vectorial Kernels: Most popular and general purpose.

Fisher Kernel:  measures the similarity of two objects on the basis of sets of measurements for each object and a statistical model.

Diffusion Kernels: Seems to be tied to the Biological field and look at the vectors in an ISOMAP.

5. How does a convolutional neural network differ from a normal neural network with dense layers?

Well, Convolutional Neural Network seems to be more useful in classifying visual data, i.e photos, video, natural language, etc.  This is due to the fact that normal neural networks typically have a full connective between one neuron in one layer to the all the neurons in the next layer, which can create overfitting.  Convolutional Neural Network instead takes advantage of the hierarchy of the data and reduces the complexity of the data to be feed int the neural network.

6. While convolutional networks use regional similarities, recurrent neural networks aim to model what?

Use sequential information, in other words, we are using inputs and outputs as independent of each other.  That way, the prediction model uses the previously feed information and the current one to make a model.  Recurrent Neural Network Work best in speech recognition since it can make a prediction as you are speaking, but based on the previous “word” spoken it may adjust itself.

7. What is an autoencoder network? What kind of learning does a autoencoder use?

Autoencoder kind of auto describes itself.  I believe this type of network is very low-level unsupervised method.  It takes the input data, takes it to the bottleneck or compressor, and then tries to have the outputs match as closely as the inputs.  Because it is unsupervised and being forced feed to the compressor it may find correlations between inputs and take advantage of it in order to compress the data.  The best way I can illustrate this is by a coin sorting machine.  X numbers of coins go in and the same X go out, but now they are sorted.