1. [18 points] Explain the key properties of the following concepts:
a) Supervised Learning
With the supervised learning concept, you will need to feed in training data to the algorithm to get a desired outcome, and this is done with labels. For example, this concept often is a classification issue of target numbers. To get the desired outcome, you will need to give a set of features which are called predicators.

2) Unsupervised Learning
Unsupervised learning is the opposite of supervised learning, in that this learning data is not labeled and the model tries to learn without a teacher. Unsupervised Learning is known for Clustering, and Anomaly Detection. Hierarchical Clustering is often used where one is subdivided into smaller groups. Visualization algorithms is also another example where you feed the model various complex unlabeled data is it will output a 2D or 3D representation of the data plotted. Another aspect of unsupervised learning is dimensionality reduction where the idea is to simplify the data without losing information. This often is done by merging several correlated features into one. Anomaly detection is also another concept of unsupervised learning, where unusual credit card transitions can be found, catching manufacturing defects, or automatically removing outliers from a dataset before feeding it to another learning algorithm. Novelty detection is another key concept of unsupervised learning, where the difference is that novelty detection algorithm expects to see only normal data during training, while anomaly detection algorithms are usually more tolerant, and often perform well even with a small percentage of outliers. Lastly the association rule learning is another concept where the goal is to dig into a large amount of data and discover interesting relations between attributes.

3) Online Learning
In online learning the system is trained incrementally by feeding data instances sequentially, individually or by mini batches. Each learning step is fast cheap, and the system can learn very quickly. This type of learning is ideal for systems that receive continuous flow, for example the stock market, and that need to change rapidly. It is also a good option if you work with limited computing resources. This is because when the system learned about the new data instance, it does not need them anymore, and they can be thrown out, which can save a huge amount of space. This learning algorithm can also be used to train systems with a vast amount of data, that generally cannot fit in one machine's main memory. This is done by loading part of the data, running the training step on that data, and repeating the process until the data is completed. Another important concept in online learning is the learning rate, which is how fast the system should adapt to changing data. If you have a high learning rate, the system will adapt to new information quickly, but it also will forget old data quickly. On the flip side if you have a low learning rate, it will learn slowly, but will be less sensitive to noise in the new data or sequences of nonrepresentative data points. Online learning does include a big challenge with if there is bad data fed to the system. If this happens, the system's performance will gradually decline. In order to reduce the risk of that, one will need to monitor the system, and promptly switch learning off, if you detect a drop in performance.

4) Batch Learning
Batch learning is the model where the system is incapable of learning incrementally. The model is trained using all the available data. This type of learning can take a lot of time and can require a large amount of computing power, and is usually done offline. With Batch learning, the system is trained, and it is launched into prod without learning anything further. If the need arises to add new data to the system, the steps must be repeated. The system must be trained on the new version of the data from the full dataset, which means new data acquired plus the old data, and the old system has to be stopped and replaced with the new systems. This type of learning can be very costly since it can take often a few hours, and training using the full data set can take many hours and due to the time constraint, it is recommended to retrain the system only every 24 hours.

5) Model-based Learning,
Model-based learning is a way to generalize to build a model with examples to make predictions. One aspect of this learning is model selections, where selected a linear model of life satisfaction with just one attribute. A simple linear model needs two model parameters, $\theta_0$ and $\theta_1$. These parameters can be tweaked, and you can represent any model as a linear function. You would also need to set a performance measure, which can be done by either a utility function or by a cost function. The utility function measures how good your model is, and the cost function will measure the distance between the linear model's predictions and the training examples. Here the objective is to minimize the distance. The Linear Regression algorithm is then used, where the training examples are fed, and it then will find the parameters that make the linear model fit bets to the provided data. This is what happens when we talk about training the model. After that step you have done all the preparations and are able to run the model and make your predictions.

6) Instance-based Learning.
Instance base learning is the idea that the system learns the examples by heart, then generalizes the new case by comparing them to the learned examples using a similarity measure. One of the examples that the book used was the spam filter. For the spam filter, your email will compare two emails based on similarity and if one was deemed spam, it looks for similarity queues to determine if another is spam as well.