This video is about gradient descent which is one way that neural networks learn. Training a network works by giving known data and outputs and using it to train. In the identifying hand written number examples the first run has useless data since all of the tuning values where randomly set. In order to improve the output a cost function is used that compares the AI output with the desired output. The output of this cost function is small when the network is close to correct but large when the network is off. After running a bunch of training examples, the cost of each example is compared. This comparison results in a slope that allows us to shift the values in a direction towards the local minimum. This is the biggest issue for accuracy is that the initial random values may not result in the global minimum being found however the local minimum should be good enough. The gradient decent refers to traversing this function and the size of the step taken on this function is dependent on how steep the slope is. By using just these methods the network gets up to about 96% accuracy struggling mostly with numbers that do not look that close to what we expect that number to look like. By changing the structure of the hidden layers and a few other adjustments this network can reach 98% accuracy. In the first video for this example the network was expected to break numbers into segments with the second layer identifying patterns and the end output piecing together these patterns to guess a number. However due to the constraints of the training and cost function the network isn’t that advanced; in fact given an image of digital noise it will output an answer with the same level of confidence that it would with an image of a number.