**Note:** You should google every question for better understanding. We will also discuss the confusion related to topics in the Question/Answer session. (if required by the class. Please ask your CR)

Video 6

Q1. What is one hot encoding? How it is helpful/benefit?  
Ans: It is the conversion of probability scores of input data into vector form where highest scored class becomes 1 and rest are set to zero. It is useful for computer in not to get confused while classification of input data. One-hot means all ‘0’ except single ‘1’.

Q.2What is scaling? Discuss advantage and disadvantage of scaling  
Ans: Scaling is the technique of standardizing the features of data within a defined range. It tells us that how a particular feature is comparatively significant as compared to other features.  
Advantages:  
1. It makes your training faster.   
2. It prevents you from getting stuck in local optima.   
3. It gives you a better error surface shape.  
Disadvantages:  
1. It can Destroy the much-desired SPARSITY of Input Data.   
2. You may forget to look for the true coefficients in regressions

Video 7

Q1. What is logistic classifier?  
Ans. The logistic classifier is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick

Q2. What is entropy? Discuss the effect of higher/lower entropy?  
Ans. Entropy is a measure of the randomness in the information being processed. The higher the entropy, the harder it is to draw any conclusions from that information

Video 8

Q.1 What is meant by optimization? Discuss the advantage of optimization.  
Ans. Optimization is the process that is associated with the improvement of a system by minimizing loss function and cost function and giving the best possible solution.

Q2. What is loss function? Discuss the roll of loss function.  
Ans. A loss function or cost function is a function that maps an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event

Q3What is meant by gradient descent?  
Ans. Gradient descent is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function. To find a local minimum of a function using gradient descent, we take steps proportional to the negative of the gradient of the function at the current point

Video 9

Q1. What problems we faced with gradient descent?

Ans.   
1. non-convex optimization problem, gradient descent only works for problems which have a well-defined convex optimization problem.  
2. If the execution is not done properly while using gradient descent, it may lead to problems like vanishing gradient or exploding gradient problems. non-convex optimization problem.

Q2. What do we mean by the term learning rate?  
Ans. The learning rate is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum of a loss function

Video 10

Q1. Discuss different famous algorithms of deep learning.  
Ans.  
1. Convolutional Neural Network (CNN)   
2. Recurrent Neural Networks (RNNs)   
3. Long Short-Term Memory Networks (LSTMs)   
4. Stacked Auto-Encoders   
5. Multilayer Perceptron Neural Network (MLPNN)   
6. Deep Boltzmann Machine (DBM)   
7. Deep Belief Networks (DBN)

Q2. How can we increase the accuracy of deep learning algorithm? (Answer in term of mean and variance).  
Ans. We can do that by keeping points in mind such as.  
1. Mean should be zero.  
2. Variance should be minimum.

Q3. What do we mean by momentum in term of deep learning algorithm?  
Ans. The basic idea of momentum in ML is to increase the speed of training. This concept is one of those small bells and whistles that you think is not as important but turns out to be a real time saver and makes things go a lot smoother

Video 11

Q1. What is learning rate decay?  
Ans. Learning rate decay describes that how learning rate varies with time as per training iteration. Learning rate decay decrease the learning rate linearly from the large initial value to the small value.

Q2. How can we convert gradient descent into stochastic gradient descent?  
Ans. We can convert gradient descent into stochastic gradient descent by fine tuning the following parameters.   
->Weight Initialization.  
->Momentum.  
->Initial Learning Rate.  
->Learning Rate Decay.

Q3. Write advantage and disadvantage of high learning rate.  
Advantages:  
1. Model will learn faster.  
2. Low computation power is required.  
Disadvantages:  
1. It can cause undesirable divergent behavior in loss function

Q4. What is adagrad?  
Ans. Adaptive Gradient is an algorithm that optimizes three of the five parameters i.e. Initial learning rate, momentum, learning rate decay. It optimizes those parameters better whose learning rate are very sparse.