Basics of Natural Language Processing(NLP):

1. 1.Explain about Bag of Words?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/bag-of-words-bow/)
2. Explain about Text Preprocessing: Stemming, Stop-word removal, Tokenization, Lemmatization. (https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/text-preprocessing-stemming-stop-word-removal-tokenization-lemmatization/)
3. Explain about uni-gram, bi-gram, n-grams.?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/uni-gram-bi-gram-n-grams/)
4. What is tf-idf (term frequency- inverse document frequency)(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/tf-idf-term-frequency-inverse-document-frequency/)
5. Why use log in IDF?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/why-use-log-in-idf/)
6. Explain about Word2Vec.?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/word2vec/)
7. Explain about Avg-Word2Vec, tf-idf weighted Word2Vec?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/avg-word2vec-tf-idf-weighted-word2vec/)
8. Explain about Multi-Layered Perceptron (MLP)?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/multi-layered-perceptron-mlp/)
9. How to train a single-neuron model?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/training-a-single-neuron-model/)
10. How to Train an MLP using Chain rule ?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/training-an-mlp-2/)
11. How to Train an MLP using Memoization?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/training-an-mlp/)
13. Explain about Backpropagation algorithm?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/backpropagation/)
14. Describe about Vanishing and Exploding Gradient problem?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/vanishing-gradient-problem-2/)
15. Explain about Bias-Variance tradeoff in neural Networks?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/bias-variance-tradeoff-23/)
16.

Deep Learning:

1. What is sampled softmax?
2. Why is it difficult to train a RNN with SGD?
3. How do you tackle the problem of exploding gradients? (By gradient clipping)
4. What is the problem of vanishing gradients? (RNN doesn't tend to remember much things from the past)
5. How do you tackle the problem of vanishing gradients? (By using LSTM)
6. Explain the memory cell of a LSTM. (LSTM allows forgetting of data and using long memory when appropriate.)
7. What type of regularization do one use in LSTM?
8. What is Beam Search?
9. How to automatically caption an image? (CNN + LSTM)
10. What is the mathematical motivation of Deep Learning as opposed to standard Machine Learning techniques?
11. In standard Machine Learning vs. Deep Learning, how is the order of number of samples related to the order of regions that can be recognized in the function space?
12. What are the reasons for choosing a deep model as opposed to shallow model? (1. Number of regions O(2^k) vs O(k) where k is the number of training examples 2. # linear regions carved out in the function space depends exponentially on the depth.)
13. How Deep Learning tackles the curse of dimensionality?
14. Why do RNNs have a tendency to suffer from exploding/vanishing gradient? How to prevent this? (Talk about LSTM cell which helps the gradient from vanishing, but make sure you know why it does so. Talk about gradient clipping, and discuss whether to clip the gradient element wise, or clip the norm of the gradient.)
15. What is the problem with sigmoid during backpropagation? (Very small, between 0.25 and zero.)
16. What is transfer learning?
17. Write the equation describing a dynamical system. Can you unfold it? Now, can you use this to describe a RNN? (include hidden, input, output, etc.)
18. What determines the size of an unfolded graph?
19. What are the advantages of an unfolded graph? (arbitrary sequence length, parameter sharing, and illustrate information flow during forward and backward pass)
20. What does the output of the hidden layer of a RNN at any arbitrary time t represent?
21. Are the output of hidden layers of RNNs lossless? If not, why?
22. RNNs are used for various tasks. From a RNNs point of view, what tasks are more demanding than others?
23. Discuss some examples of important design patterns of classical RNNs.
24. Write the equations for a classical RNN where hidden layer has recurrence. How would you define the loss in this case? What problems you might face while training it? (Discuss runtime)
25. What is backpropagation through time? (BPTT)
26. Consider a RNN that has only output to hidden layer recurrence. What are its advantages or disadvantages compared to a RNNhaving only hidden to hidden recurrence?

27. What is Teacher forcing? Compare and contrast with BPTT.
28. What is the disadvantage of using a strict teacher forcing technique? How to solve this?
29. Explain the vanishing/exploding gradient phenomenon for recurrent neural networks. (use scalar and vector input scenarios)
30. Why don't we see the vanishing/exploding gradient phenomenon in feedforward networks? (weights are different in different layers – Random block intialization paper)
31. What is the key difference in architecture of LSTMs/GRUs compared to traditional RNNs? (Additive update instead of multiplicative)
32. What is the difference between LSTM and GRU?
33. Explain Gradient Clipping.
34. Adam and RMSProp adjust the size of gradients based on previously seen gradients. Do they inherently perform gradient clipping? If no, why?
35. Discuss RNNs in the context of Bayesian Machine Learning.
36. Can we do Batch Normalization in RNNs? If not, what is the alternative? (BNorm would need future data; Layer Norm)
37. What is representation learning? Why is it useful? (for a particular architecture, for other tasks, etc.)
38. What is the relation between Representation Learning and Deep Learning?
39. What is one-shot and zero-shot learning (Google's NMT)? Give examples.
40. What trade offs does representation learning have to consider?
41. What is greedy layer-wise unsupervised pretraining (GLUP)? Why greedy? Why layer-wise? Why unsupervised? Why pretraining?
42. What were/are the purposes of the above technique? (deep learning problem and initialization)
43. Why does unsupervised pretraining work?
44. When does unsupervised training work? Under which circumstances?
45. Why might unsupervised pretraining act as a regularizer?
46. What is the disadvantage of unsupervised pretraining compared to other forms of unsupervised learning?
47. How do you control the regularizing effect of unsupervised pre-training?
48. How to select the hyperparameters of each stage of GLUP?

External sources

https://www.analyticsvidhya.com/blog/2017/01/must-know-questions-deep-learning/