1. What are representation and representation learning meaning?
2. Why do we need to perform representation learning in many machine learning applications?
3. What’s the relationship between AI, machine learning and deep learning?
4. What are L0 norm, L1 norm, and L2 norm?
5. What’s the definition of the Frobenius norm of matrices?
6. What’s the chain rule of conditional probabilities?
7. Please write out the probabilistic density function of multivariate Gaussian.
8. How to compute the Kullback-Leibler (KL) divergence of two distributions?
9. What’s the condition number?
10. What are the local minimum and saddle point?
11. How to evaluate the abilities of a machine learning algorithms?
12. What are overfitting and underfitting?
13. What is the curse of dimensionality?
14. For a regression problem with real outputs, what kind of output units should we choose? Why?
15. Why do we generally use ReLU as the activation function other than sigmoid/tanh in CNNs?
16. How to use the back propagation algorithm to optimize deep neural networks?
17. Why do we apply regularization on deep learning models?
18. Please introduce at least 3 norm based regularization methods.
19. Why do we consider multi-task learning as a regularization method?
20. How to perform dropout regularization?