Twitter Thread by Santiago



Santiago
@svpino

20 machine learning questions that will make you think.

(Cool questions. Not the regular, introductory stuff that you find everywhere.)

■■

1. Why is it important to introduce non-linearities in a neural network?

2. What are the differences between a multi-class classification problem and a multi-label classification problem?

3. Why does the use of Dropout work as a regularizer?

■

4. Why you shouldn't use a softmax output activation function in a multi-label classification problem when using a one-hot-encoded target?

5. Does the use of Dropout in your model slow down or speed up the training process? Why?

■

6. In a Linear or Logistic Regression problem, do all Gradient Descent algorithms lead to the same model, provided you let them run long enough?

7. Explain the difference between Batch Gradient Descent, Stochastic Gradient Descent, and Mini-batch Gradient Descent.

■

8. What are the advantages of Convolution Neural Networks (CNN) over a fully connected network for image classification?

9. What are the advantages of Recurrent Neural Networks (RNN) over a fully connected network when working with text data?

■

10. How do you deal with the vanishing gradient problem?

11. How do you deal with the exploding gradient problem?

12. Are feature engineering and feature extraction still needed when applying Deep Learning?

13. How does Batch Normalization help?

■

14. The universal approximation theorem shows that any function can be approximated as closely as needed using a single nonlinearity. Then why do we use more?

15. What are some of the limitations of Deep Learning?

16. Is there any value in weight initialization?

■

17. The training loss of your model is high and almost equal to the validation loss. What does it mean? What can you do?

18. Assuming you are using Batch Gradient Descent, what advantage would you get from shuffling your training dataset?

■

19. Compare the following evaluation protocols: a hold-out validation set, K-fold cross-validation, and iterated K-fold validation. When would you use each one?

20. What are the main differences between Adam and the Gradient Descent optimization algorithms?

■

I'll be posting an in-depth answer for each one of these questions over the next following days.

Stay tuned and help me spread more machine learning content to more people in the community!

Feel free to post answers for those questions that you know. Give them a try!

You'll be forced to think about them just by trying to collect your thoughts to put them here.

(There's nothing as effective to learn than interacting with others.)