

Midterm Machine Learning

Professor Francesco Zuppichini

November 7, 2017

1. What is gradient descent? State its generic update rule and show why it can be used in ML
2. What is a Perceptron? State the forward pass and update rule.
3. Compute the new weights and bias after one full epoch on a Perceptron with input $x = \begin{pmatrix} 2 \\ 10 \end{pmatrix}$, weight $w = \begin{pmatrix} 1.2 \\ -0.5 \end{pmatrix}$ and bias $b = 3$.
4. What is a Neural network? Give a high-level definition.
5. Which are the hyper-parameters in a Neural Network?
6. Which problem can a neural network solve that a perceptron can not? Give at least one example.
7. State the delta rules for the weights and bias.
8. Give a definition for the following term: hidden layer, activation function, back-propagation and regression
9. Quickly define overfitting and underfitting.
10. Write down at least two activation functions and their derivative
11. What is stochastic gradient descent and what problem can it solve?
12. Define regression problems and classifications. State the relative loss functions
13. What is a Convolutional Neural Network and for what they are used for?
14. Why is a Convolutional Layer different from a normal one?

15. Give a definition for the following terms: stride and max-pooling
16. Why RELU is used in a CNN? (write it down its equation)
17. Why the last layer is a normal fully connected layer? Why we need softmax in the end?
18. What is a Recurrent Neural Network? Sketch it down a generic example and give a high-level definition.
19. State the forward pass and the update rules for weight and bias in a RNN. Give a quick overview.
20. Vanishing Gradient is a well-know problem in RNN. Explain it.
21. Why where LSTM introduced?
22. Draw a LSTM-cell and describe it in deep.
23. Support Vector Machine are another supervise learning algorithm. Give generic definition.
24. The simplest SVM is just a binary linear classifier. Sketch an example and give an explanation. Use equations to support your words.
25. Sketch an example of non-linear separable data before and after the mapping to the feature space. Also, what is it?
26. What is the kernel? Why it is so important? Do you know any trick that can be done with it?
27. Explain how the dual problem is stated, what we want to maximise/minimise and how Quadratic Programming can help do that. Do not focus too much on the math