ECE542 - Quiz 4

1. [2 pts] Based on the universal approximation theorem, how many hidden layers are needed in order to have the set of functions that can be represented by a neural network to be asymptotically dense in $C(\mathbb{R})$?

2. [2 pts] Remember that: "A family of real functions S **separates point** in [0,1] if for every $x, y \in [0,1]$, $x \neq y$, there exists a function $g \in S$ such that $g(x) \neq g(y)$." Does the set of polynomials separate point in [0,1]? If so, provide a function that separates points.

3. [1 pts] Fill-in the blank: *Adaptive learning rate methods (e.g., RMSProp and Adam) often scale the learning rate by accumulating the magnitude of the* _____.

4. [3 pts] For batch normalization, we aim to have (ignoring the effect of γ and β):

   a) A mean of the for the variable a_i^l in layer l (i.e., $E[a_i^l]$ for $i = 1, \dots, m$) equals to?

   b) A variance for the variable (i.e., $var(a_i^l)$) equals to?

   c) A covariance c_{ij} between values at different neurons (i.e., $cov(a_i^l, a_j^l)$) equals to?

Answer

1. 1
2. Yes, g(x) = x
3. Gradient
4. a) 0
   b) 1
   c) [0,1]