Written assignment

一、Lemma 3.1

1. Background of Lemma:

   (1) Shallow tanh neural network: shallow neural network means the neural network only has 2 layers. Tanh means the activation function is $\sigma(x) = tanh(x)$.

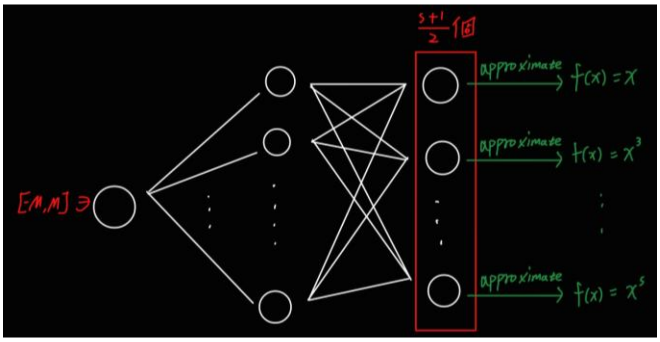
   (2) Width of neural network: the width means the greatest number of neurons of all layers.

   (3) This lemma can be applied to different activation functions like sigmoid or logistic function.

2. Lemma statement:

   given a non-negative integer k_0 and an odd number s, for any $\varepsilon > 0$, there exists a shallow tanh neural network $\psi_{s,\varepsilon} : [-M, M] \to \mathbb{R}^{\frac{s+1}{2}}$ of width $\frac{s+1}{2}$ such that when $p \leq s$ is an odd number, the error between $f_p(x) = x^p$ and $\left(\frac{p+1}{2}\right)$-th element of $\psi_{s,\varepsilon}$ smaller than ε. In other words, we can approximate f_p accurate as we want by shallow tanh neural network because s can be chosen arbitrary, where p is an odd number.

   Moreover, the weights of $\psi_{s,\varepsilon}$ will be limited corresponding to ε and s.



3. Example:

   If $s = 5$, then we can use $(\psi_{5,\varepsilon})_1$, $(\psi_{5,\varepsilon})_2$, $(\psi_{5,\varepsilon})_3$ to approximate x, x^3, x^5 for accuracy to ε.

二、Lemma 3.2

The background of this lemma is the same as lemma 3.1. And it is the extend of lemma 3.1. Lemma 3.1 only said that we can approximate the monomials of odd degree. 3.2 said that for all monomials can be approximated by shallow neural network and the weights will be limited corresponding to ε and s.

In this lemma, there exists a shallow tanh neural network $\psi_{s,\varepsilon}$: $[-M, M] \to \mathbb{R}^s$ of width $\frac{3s+3}{2}$ such that the error between $f_p(x) = x^p$ and p-th element of $\psi_{s,\varepsilon}$ smaller than ε, where s is an odd number and $p \leq s$. For example, If s=5, then we can use $(\psi_{5,\varepsilon})_1$, $(\psi_{5,\varepsilon})_2$, $(\psi_{5,\varepsilon})_3$, $(\psi_{5,\varepsilon})_4$, $(\psi_{5,\varepsilon})_5$ to approximate x, x^2, x^3, x^4, x^5 for accuracy to ε.