2023/01/20

Exercise 3

Show that a multi-layer neural network with linear activation function $s(x) = x$ is equivalent to a single layer linear network. Assume that in each layer the inputs follow a Normal distribution with mean zero and small variance, *i.e.* $\sigma \ll 1$. For which of the activation functions $s(x) = 1/(1+\exp(-x))$, $s(x) = \tanh(x)$, $s(x) = \mathrm{relu}(x)$ and $s(x) = \mathrm{selu}(x)$ is a deep network equivalent to a linear network for the given distribution ? The selu function is given by:

$$\mathrm{selu}(x) = \begin{cases} \lambda x \text{ if } x > 0 \\ \alpha(\exp(x) - 1) \text{ otherwise} \end{cases}, \tag{4}$$

where $\lambda \approx 1.0507$ and $\alpha \approx 1.75814$. (*Hint:* consider the case $\sigma \to 0$ using a Taylor expansion around 0.)

- Deep Linear = single linear

  MLP (with L layers) has output: $y = W^L \cdot W^{L-1} \cdot \ldots W^2 \cdot W^1 \cdot x$

  we can just rewrite $W^L \cdot W^{L-1} \ldots W^2 \cdot W^1 = W_{eff}$ $\longrightarrow y = W_{eff}\, x$

  (single neuron!)

- for which activation functions is a deep approximately linear when $x \sim N(0, \sigma^2)$ with $\sigma^2 \ll 1$?

1) Sigmoid $s(x) = \frac{1}{1+e^{-x}}$

   Taylor expansion in $\emptyset$: $s(x) \approx \frac{1}{2} + \frac{x}{4} - \frac{x^3}{48} + \ldots$

   for small σ $s(x) \approx \frac{1}{2} + \frac{x}{4}$ Not linear: constant bias breaks linearity over multiple layers ✗

2) $\tanh(x)$

   Taylor expansion $s(x) = x - \frac{x^3}{3} + \frac{2x^5}{15} \ldots$

   for small σ $s(x) = x$ Linear ✓

3) $\mathrm{ReLU}(x) = \max(0, x)$

   $x < 0$ → output 0

   $x > 0$ → output x

   ↳ normal distribution around $\emptyset$ means that half input are mapped into $\emptyset$

   Non linearity!

4) Selu(x) = $\begin{cases} \lambda x & x > 0 \\ \lambda\alpha(e^x - 1) & x \le 0 \end{cases}$

Taylor expansion $\begin{cases} \lambda x & x > 0 \\ \lambda\alpha(x + \frac{x^2}{2} + \dots) \approx \lambda\alpha x & x \le 0 \end{cases}$

↳ this is piecewise linear (but not globally) slope differs between $x > 0$ and $x < 0$