2022/01/18

Exercise 3

Consider the neural network $\mathcal{N}$ defined by the following composite function:

$$\mathcal{N}(\mathbf{x}) = C(ReLU(B(ReLU(A\mathbf{x})))), \tag{2}$$

where:

- $ReLU$ is applied componentwise;
- $A, B \in \mathbb{R}^{50\times 50}$;
- $C \in \mathbb{R}^{5\times 50}$.

1. For the given matrices A, B and C what are the dimensions of the input and output of the neural network $\mathcal{N}$? Can the output be seen as a vector of probabilities ?
2. Consider the softmax function
$$\text{softmax}(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{3}$$
for $i \in \{1, \dots, n\}$.
For the given matrices A, B and C what are the dimensions of the input and output of the neural network softmax($\mathcal{N}$) ? Can the output be seen as a vector of probabilities ?
3. If $\mathcal{N}$ is a classifier, for how many classes does it work?
4. Assume $\mathcal{N}$ has been trained and you want to use it to for classification on an input $\mathbf{x}$. Do you use softmax($\mathcal{N}$) or $\mathcal{N}$? Justify your answer.

$N(x) = C \cdot ReLU(B \cdot ReLU(A \cdot x))$ $\quad A \in \mathbb{R}^{50\times 50}$ $\quad B \in \mathbb{R}^{50\times 50}$ $\quad C \in \mathbb{R}^{5\times 50}$

- 3 Layer Feed Forward Neural Network

input vector $x \in \mathbb{R}^{50\times 1}$

- $Ax \in \mathbb{R}^{50}$
- $ReLU(Ax) \in \mathbb{R}^{50}$
- $B \cdot ReLU(Ax) \in \mathbb{R}^{50}$
- $ReLU(B \cdot ReLU(Ax)) \in \mathbb{R}^{50}$
- $C \cdot ReLU(\cdot) \in \mathbb{R}^{5}$

input $\mathbb{R}^{50}$, output $\mathbb{R}^{5}$

↳ We are not guaranteed to have $N(x)$ $\begin{cases} \text{non-negative} \\ \text{summing to 1} \end{cases}$, cannot be seen as probability

· If we apply softmax: $\text{softmax}(N(x))_i = \frac{e^{N(x)_i}}{\sum_{j=1}^{5} e^{N(x)_j}}$

input of softmax $\mathbb{R}^5$

output of

└ Now output can be seen as probabilities! { each component of output is in $(0,1)$; they sum to 1

→ It would be a 5-class classifier

→ Use softmax if we want probabilities

Use $\text{argmax}(N(x))$ if we are interested only in the predicted class

(in principle both can be used)