

Chapter theory 3 solution

Artem Puzanov

2016-09-27

1 Proof of universal computability

1.1 Step wise function in any direction

Intuition To prove that we can generate step-function in any direction, we will show that we can generate steps orthogonal to x_1, x_2, \dots, x_n plane such that on the one side of the plane value of activation is close to 1 and on the other it's close to 0. For 2 variables (x, y) , this translates into a x, y line, on the one side of which $\sigma(z)$ is 1, and on the other it's 0.

Mathematical notation Let's define x_1, x_2, \dots, x_n as X . We shall take arbitrary vector of coefficients $K = k_1, k_2, \dots, k_n$. Then our arbitrary plane can be written as $K \odot X = C$. This means that we have $n - 1$ free variables, and one that is a linear combination of them.

$$x_j = C/k_j + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j}$$

where K_{-j}, X_{-j} refer to vectors K and X without j -th element.

Now, what's a step outside this plane? Well, it's the parallel plane (don't call me Sirely):

$$K \odot X = C + \epsilon$$

where ϵ is the size of our step. Note that this translates into

$$x_j = \frac{C + \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j}$$

This allows us to put all those coefficients into activation function. For the statement to be true, we have to prove that these two inequalities hold:

$$\forall \delta \in (0, 0.5), \epsilon > 0, C \in R, K \in R : \exists W :$$

$$(1) \frac{1}{1 + \exp(-\sum_{i \neq j} w_i x_i - w_j (\frac{C + \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j}))} > 1 - \delta$$

$$(2) \frac{1}{1 + \exp(-\sum_{i \neq j} w_i x_i - w_j (\frac{C - \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j}))} < \delta$$

Let's simplify (example for (1)):

$$1 + \exp(-\sum_{i \neq j} w_i x_i - w_j (\frac{C + \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j})) < 1 - \delta$$

$$1 + \exp(-\sum_{i \neq j} w_i x_i - w_j (\frac{C + \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j})) < \frac{1}{1 - \delta}$$

$$\exp(-\sum_{i \neq j} w_i x_i - w_j (\frac{C + \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j})) < \frac{\delta}{1 - \delta}$$

$$\begin{aligned}
-\sum_{i \neq j} w_i x_i - w_j \left(\frac{C + \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j} \right) &< \ln\left(\frac{\delta}{1 - \delta}\right) \\
\sum_{i \neq j} w_i x_i + w_j \left(\frac{C + \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j} \right) &> \ln\left(\frac{\delta}{1 - \delta}\right)
\end{aligned}$$

Applying the same for (2) we get:

$$\sum_{i \neq j} w_i x_i + w_j \left(\frac{C - \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j} \right) < \ln\left(\frac{1 - \delta}{\delta}\right)$$

Let's define $\ln\left(\frac{\delta}{1 - \delta}\right) = A$ Than our inequalities transform into:

$$\begin{aligned}
(1) \sum_{i \neq j} w_i x_i + w_j \left(\frac{C + \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j} \right) &> A \\
(2) \sum_{i \neq j} w_i x_i + w_j \left(\frac{C - \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j} \right) &< -A \\
(2) - \sum_{i \neq j} w_i x_i - w_j \left(\frac{C - \epsilon}{k_j} + \frac{\langle K_{-j}, X_{-j} \rangle}{k_j} \right) &> A
\end{aligned}$$

Combining these two, we get:

$$w_j \epsilon / k_j > 2A$$

$$w_j > A k_j / \epsilon$$

This inequality states that for any choice of $\{\delta, \epsilon, K, C\}$ we can always find a weight W to satisfy conditions of our step function. Q.E.D.

1.2 Combine multiple steps into a circle

Intuition The idea is simple - use enough straight lines at small angles to each other, get a circle. In general sense - get enough linear planes (don't call me Shirley), get a hypersphere.

Mathematical notation None needed, nothing to prove yet

1.3 Proof of several towers

Intuition We can build many towers, the trick is to make them so that we can manipulate their height. How to achieve this? By using window of opportunity between bias and default tower height. This way our tower sticks out, yet is zero where we want it to be zero.

Mathematical notation Let's assume we used k step functions to make an approximate hypersphere, and we did it for m towers. That gives us weighted output from hidden layer in the form:

$$z = \sum_i^m h_i + (k - 0.5)h$$

where $h_i = \langle w_i, x_i \rangle$, being combination of step functions to generate the particular activated tower. Now, what we want to achieve is to have towers of different height. That can be done with adding a new multiplier to each h_i :

$$z = \sum_i^m T_i h_i + (k - 0.5)h$$

where T_i is our tool for tower height regulation. Let's say we want the tower height to be T' . Then we have to solve this equation for T_i

$$\sigma(0.5h_i T_i) \odot f(x) = T'$$

where h_i is a known constant. It's obvious there is a solution for any T' . A note: we didn't use the *tower shift* trick suggested by author, might need some thinking.

1.4 Proof for rectified linear unit

Intuition It's easy to see why RLU doesn't satisfy the universality conditions - it's not well defined: $\lim_{z \rightarrow \infty} s(z) \rightarrow \infty$. Yet there is a step, it's just not ending after step. So we need to find a way to combine several RLU's into a step.

The proper way to do it is this:

- 1) Take one RLU, turn it's linear coefficient into lava (use whatever bias you want)
- 2) Take another RLU, with the negative lava coefficient, and shifted and negative bias.
- 3) Sum them up

This will give you a zone where one unit is active, another is not. This zone goes up to a constant difference between to RLU's, effectively giving you a step function. After that - you know what to do).

So to summarize - we need one layer of [pair/triple..]wise RLU's, next one to turn them into stepwise functions, and the final one to generate towers. The only difference from well-defined neurons is necessity of using combinations of RLU's to generate step functions.