Worksheet 5: Universal Approximation

Name: Aaron Zoll
Due October 25, 2023
In this worksheet, we will constructively tie up loose ends from universal approximation theorem, to the point of rendering lemmas 11.1 and 11.2 from the notes almost superfluous.

1. Let $\mathrm{K} \subset \mathbb{R}$ be a compact interval, and $\mathcal{O} = \{\mathrm{U}_\alpha\}_{\alpha \in \mathrm{A}}$ an open cover. Show that there is refinement $\mathcal{O}' = \{\mathrm{I}_\beta\}_{\beta \in \mathrm{B}}$ consisting of open intervals I_β-i.e. for each $\beta \in \mathrm{B}$, there is $\alpha \in A$ for which $\mathrm{I}_\beta \subset \mathrm{U}_\alpha$-with the property that each $x \in \mathrm{K}$ is contained in at most two I_β s from $\mathcal{O}'$. (It may be helpful, though not necessary, to remind yourself of Heine-Borel.)
   Solution: Since K is compact, there exists a smallest finite subcover $\mathcal{O}_{fin} = \{U_k\}_{k=1}^n$. This isn't necessary, but this makes it easier for later problems to only have a finite number of open sets to deal with. Then, because K is closed, it contains all its limit points. That is, for any $x_n \in U_\alpha$ where $x_n \to x$, it must also hold that $x \in K$. It may not hold that $x \in U_\alpha$, but regardless, $x \in U_{\alpha'}$. Then since $U_{\alpha'}$ is open there is some $\epsilon > 0$ where $B_\epsilon(x) := \{y : \|y - x\| < \epsilon\}$ is fully contained in $U_{\alpha'}$. However, by the definition of limits, there is also (an infinite in fact) number of elements $x_n \in U_{\alpha'}$ for all $n > N_\epsilon$

   This is all to say that $U_\alpha \cap U_{\alpha'} \neq \emptyset$. We can then take $(a_{\alpha_i}, b_{\alpha_i})$ to be the largest open interval contained in each of the $i \in I_\alpha$ connected components of each U_α. Technically, for each connected component of U_α. Pick an $x \in U_{\alpha_i}$ and define $a_{\alpha_i} := \inf_{(a,x) \subset \alpha_i}\{a \in U_{\alpha_i}\}$ and similarly, $b_{\alpha_i} := \sup_{(a,x) \subset \alpha_i}\{b \in U_{\alpha_i}\}$. This will in fact cover all of K by the construction outlined above, and each is fully contained in the respected U_{α_i}. Then since $\mathcal{O}_{fin}$ is the smallest finite subcover, we can force these intervals to only overlap at most twice. That is, if there is a circumstance where x lives in there of these intervals, we would have failed one of these extreme conditions (either for the minimality of the finite subcover, the inf of a_{α_i} or the sup of b_{α_i}.

2. Let $\mathrm{I}_0 \cap \mathrm{I}_1 \neq \varnothing$ be two intervals, with $\mathrm{I}_0 = (-\infty, \mathrm{b}_0), \mathrm{I}_1 = (\mathrm{a}_1, \infty)$ with $\mathrm{a}_1 < \mathrm{b}_0$. Explicitly detail how to construct sigmoidals $\sigma_0(t) := \frac{1}{1+e^{-w_0(t-b_0)}}$ and $\sigma_1(t) := \frac{1}{1+e^{-w_1(t-b_1)}}$ so that (a) $\sigma_0 + \sigma_1 \equiv 1$ and (b) $\sigma_0^{-1}(0, \varepsilon/2) \subset \mathrm{I}_1 \backslash \mathrm{I}_0$ and $\sigma_0^{-1}(1 - \varepsilon/2, 1) \subset \mathrm{I}_0 \backslash \mathrm{I}_1$.

   In showing the (a), make sure to (also) express σ_1 in terms of w_0 and b_0. State and conclude the analog of (b) for σ_1.

Solution: I rewrote the functions slightly to make the computation easier. For $\sigma_1(t)$ take $b_1 = \frac{a_0+b_0}{2}$ and take $w_1 = \frac{2}{b-a}\ln(\frac{\epsilon}{2-\epsilon})$. Then $\sigma_0(t) := 1 - \sigma_1(t)$ is still a sigmoidal because for any w_1 and b_1 it would hold that

$$\begin{aligned}
1 - \frac{1}{1+e^{-wt+b}} &= \frac{1+e^{-wt+b}}{1+e^{-wt+b}} - \frac{1}{1+e^{-wt+b}} \\
&= \frac{e^{-wt+b}}{1+e^{-wt+b}} \\
&= \frac{1}{\frac{1}{e^{-wt+b}} + 1} \\
&= \frac{1}{1+e^{wt-b}}
\end{aligned}$$

So it is still sigmoidal in this sense. And it isn't hard to see that $\sigma_1^{-1}(0, \epsilon/2)$ as we can just check what happens when we plug in a_0 and b_0 since this function is increasing (which is why I am checking $\sigma_1(t)$ not $\sigma_0(t)$).

$$\begin{aligned}\sigma_1(a_0) &= \frac{1}{1+e^{\frac{2}{b-a}\ln(\frac{\epsilon}{2-\epsilon})(a-\frac{a+b}{2})}} \\ &= \frac{1}{1+e^{\ln(\frac{\epsilon}{2-\epsilon})\frac{2}{b--a}(\frac{a-b}{2})}} \\ &= \frac{1}{1+\frac{2-\epsilon}{\epsilon}} \\ &= \frac{\epsilon}{2} \\ &= \epsilon/2\end{aligned}$$

Therefore, for all $x < a$ we have $\sigma_1(x) < \epsilon/2$ so $\sigma_1^{-1}(0, 1-\epsilon/2) \subset I_I \backslash I_0 = (-\infty, b_0)$
Similarly,

$$\begin{aligned}\sigma_1(b_0) &= \frac{1}{1+e^{\frac{2}{b-a}\ln(\frac{\epsilon}{2-\epsilon})(b-\frac{a+b}{2})}} \\ &= \frac{1}{1+e^{\ln(\frac{\epsilon}{2-\epsilon})\frac{2}{b-a}(\frac{b-a}{2})}} \\ &= \frac{1}{1+\frac{\epsilon}{2-\epsilon}} \\ &= \frac{2-\epsilon}{2-\epsilon+\epsilon} \\ &= 1-\epsilon/2\end{aligned}$$

3. Using the preceding problem, explain how to construct smooth step function σ_0' and σ_1' (not derivatives) satisfying (a) $\sigma_0' + \sigma_1' \equiv 1$ and (b) $\sigma_0'^{-1}(\mathrm{I}_0 \backslash \mathrm{I}_1) \equiv 1$ and $\sigma_0'^{-1}(\mathrm{I}_1 \backslash \mathrm{I}_0) \equiv 0$.

You may use the fact that any open cover of $\mathbb{R}$ has a partition of unity subordinate to it.
Solution: We can construct this the same way as part 2 using the start function of

$$s(x) = \begin{cases} 0 \ : \ x \leq 0 \\ e^{-1/x} \ : \ x > 0 \end{cases}$$

We then take some trick with something like $\frac{s(x)}{1-s(x)}$ and add different weights, but we can in fact just use the property that any open cover of $\mathbb{R}$ has a PO1 subordinate, and since $I_0 \cup I_1 = \mathbb{R}$, we can just define σ_0' and σ_1' as being subordinate to these sets.

4. For a given open cover $\mathcal{O}$ of intervals as in 1. above, explain how you would construct a partition of unity $\{\rho_\beta\}_{\beta\in\mathrm{B}}$ subordinate to $\mathcal{O}$ such that to each ρ_β there is sigmoidal $\sigma_\beta := \frac{1}{1+e^{-w_\beta(\cdot)+b_\beta}}$ for which $\|\rho_\beta - \sigma_\beta\|_\infty < \varepsilon$.
   Solution: I believe we ended up with this question actually asking about the intervals as in 2, not one. This then allows use to both have infinite supports over our open cover, and thus we can actually get sigmoidals that approximate. Having only 2 intervals as well also lets us use just sigmoidals and no needs for the span of sigmoidals.

   Using parts 2 and 3, we have that outside of $I_0 \cap I_1$, that for (WLOG) the increasing functions ρ_0 and σ_0, it holds that $\|\rho_0 - \sigma_0\|_\infty \leq |\rho_0(x) - \sigma_0(x)|$ for any x, but we just showed that this is bounded by $\epsilon/2$.

   By actually constructing the smooth step functions with $s(x)$ as outlined, it will hold that they both intersect at $\frac{a+b}{2}$ and we can show via calculation I hope I don't need to actually do that they differ by at most $\epsilon/2 + \epsilon/2 = \epsilon$. When we have more intervals, we just keep repeating this process by looking at the overlap, we may need to subtract some "same-weight, different bias" sigmoidals to get the bump functions, but these will still approximate the PO1. The only difference, is we may need to change the ϵ in the construction to and ϵ/n for the number of overlaps, because when we apply the triangle inequality mutliple times, we may need to be cautious, but the finiteness of all this will still allow for expressability.

5. Conclude: for any open cover $\mathcal{O}$ of K, there is partition of unity $\{\rho_\beta\}_{\beta\in B}$ subordinate to $\mathcal{O}$ which the class of sigmoidals $\Sigma := \left\{\sigma_{w,b} = 1/\left(1+e^{-w(\cdot)+b}\right) : w, b \in \mathbb{R}\right\}$ expresses.

Solution: In this part, we take an open cover, refine to a finite subcover, refine to intervals that overlap at most twice, and then create our bump/sigmoidal functions that approximate every PO1 restricted the the intervals as defined above. We then take $\epsilon' := \epsilon/(2N)$ where N is the size of the finite number of intervals that cover K, and we should be able to get the infinity norm bound that we want.