Computational Thinking
Sample Solutions to Exercise 11

1 Limitations of Neural Networks

A neural network can in theory approximate any continuous function given a sufficiently large number of hidden nodes. Therefore, only c) and e) cannot be represented, as those functions are not continuous.

2 An Ill-Designed Network

a) \( \hat{f}(x|a, b) = 1 \cdot \tanh(100 \cdot 0.9) = 1 \) (given numerical precision)

b) \( \frac{dL}{db} = \frac{dL}{d\hat{f}} \cdot \frac{d\hat{f}}{d \hat{f}} = (f(x) + \hat{f}(x|a, b)) \cdot \tanh(ax) = 0.1 \cdot \tanh(90) = 0.1 \)

c) \[
\frac{dL}{db} = \frac{dL}{d\hat{f}} \cdot \frac{d\hat{f}}{d \tanh(ax)} \cdot \frac{d \tanh(ax)}{dx} \\
= (f(x) + \hat{f}(x|a, b)) \cdot b \cdot (1 - \tanh^2(ax)) \cdot x \\
= 0.0 \text{ (since } 1 - \tanh^2(90) = 0). \]

d) \( a_{\text{new}} = a, b_{\text{new}} = b - 0.1 \cdot \frac{dL}{db} = 0.99. \) The weight \( a \) which causes the issue did not get any update due to a vanishing gradient, which causes the problem to persist for further updates.

e) If we do the same calculations for \( x = 0.9 \) again we find that \( \frac{dL}{da} \approx 3099.56. \) This yields \( a_{\text{new}} = a - \alpha \frac{dL}{da} \approx -308.956 \) and following updates will again have the vanishing gradient problem. The first update suffers from what is called an exploding gradient here.

[Bonus] The hyperbolic tangent is close to linear around the origin, a decent approximation would therefore be given by \( 0 < a << 1 \) and \( b = 1/a. \)

3 Gradient Descent with Momentum

a) \( \beta = 0 \)

b) Roughly at the same point where the light green cross is, as the loss surface is flat which leads to a gradient close to zero.

c) The update is much bigger into the direction of the global optimum as \( m_{\text{w}} \) is dominated by the bigger gradient from the preceding step.
d) In the global optimum.

e) The large gradients in the first few iterations might dominate \( n_w \) and drive the optimization across the global optimum up the hill into the local optimum on the right.