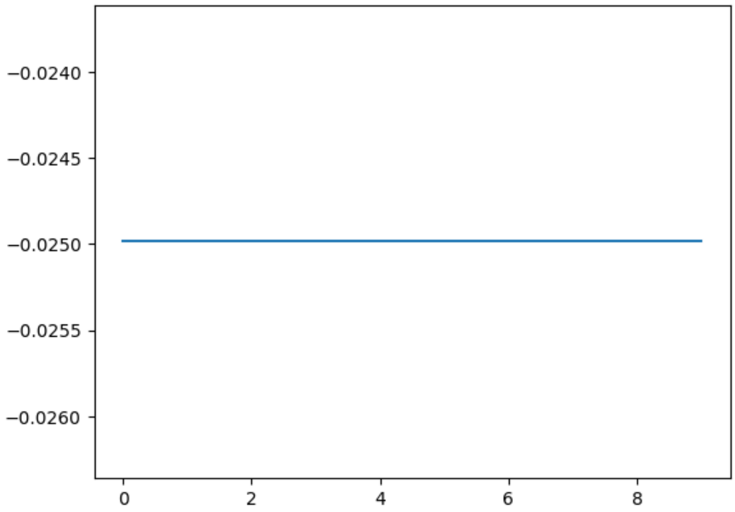
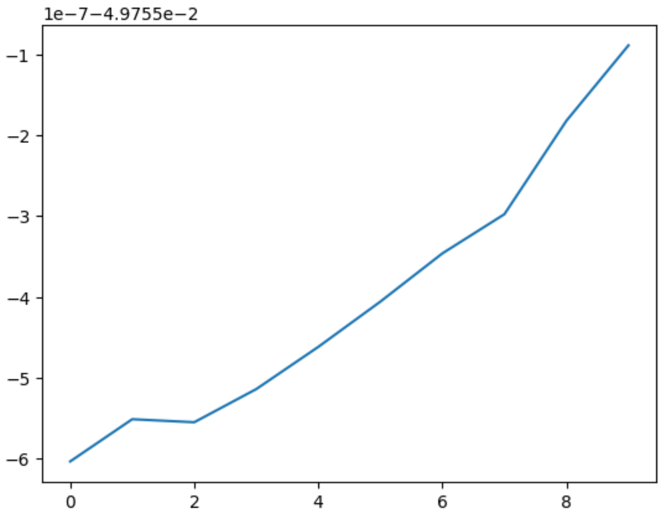
1. **Execute o código anexado nessa atividade. Corrija o problema do código e crie uma estratégia para avaliar a evolução dos pesos na primeira camada da rede (camada mais baixa) após a correção. Descreva o problema encontrado e apresente um gráfico com a evolução dos pesos.**

Evolução do peso 1 da camada 1 antes da troca da função de ativação.



Evolução do peso após a troca da função de ativação.



O problema do código é a função de ativação de cada camada. Com o uso da função sigmoid, é provável que os gradientes não são suficientemente grandes para causar alterações aos pesos das primeiras camadas. Com a alteração para a função de ativação ReLU, os pesos começam a atualizar, conforme as figuras apresentadas.

2. **Seguindo o exemplo da aula, apresente o cálculo do gradiente para o peso w7 (figura abaixo). Assuma as mesmas funções de erro e de ativação do exemplo apresentado nos slides.**

$$E_T = \frac{1}{2}(y_1 - g_{o1})^2 + \frac{1}{2}(y_2 - g_{o2})^2 \qquad u_{o2} = g(h_1)\cdot w_7 + g(h_2)\, w_8 + b_2$$

w_7

$$y_1 = 0{,}01 \qquad y_2 = 0{,}99 \qquad g(o_2) = \frac{1}{1+e^{-u_{o2}}}$$

$$\frac{\partial E_T}{\partial w_7} = \frac{\partial E_T}{\partial g_{o2}} \cdot \frac{\partial g_{o2}}{\partial w_7} = \frac{\partial E_T}{\partial g_{o2}} \cdot \frac{\partial g_{o2}}{\partial u_{o2}} \cdot \frac{\partial u_{o2}}{\partial w_7}$$

$$\frac{\partial u_{o2}}{\partial w_7} = \frac{\partial (g(h_1)\cdot w_7)}{\partial w_7} + \cancel{\frac{\partial (g(h_2)\cdot w_8)}{\partial w_7}}^{0} + \cancel{\frac{\partial (b_2)}{\partial w_7}}^{0}$$

$$= g(h_1) = 0{,}59326992$$

$$\frac{\partial g_{o2}}{\partial u_{o2}} = g_{o2}\cdot(1-g_{o2}) = 0{,}772928465\,(1 - 0{,}772928465)$$

$$= 0{,}175510053$$

$$\frac{\partial E_T}{\partial g_{o2}} = \cancel{\frac{\partial \left(\frac{1}{2}(y_1 - g_{o1})^2\right)}{\partial g_{o2}}}^{0} + \frac{\partial \left(\frac{1}{2}(y_2 - g_{o2})^2\right)}{\partial g_{o2}} \qquad a = y_2 - g_{o2}$$

$$= \frac{1}{2}\frac{\partial a^2}{\partial g_{o2}} = \frac{1}{2}\cdot 2 \cdot a \cdot \frac{\partial a}{\partial g_{o2}} = (y_2 - g_{o2})\left[\frac{\partial (y_2)}{\partial g_{o2}} + \frac{\partial (-g_{o2})}{\partial g_{o2}}\right]$$

$$= (y_2 - g_{o2}) \cdot (-1) = g_{o2} - y_2 = 0{,}772928465 - 0{,}99$$

$$= -0{,}217071535$$

$$\frac{\partial E_T}{\partial w_7} = -0{,}217071535 \cdot 0{,}175510053 \cdot 0{,}59326992 = -0{,}0226025378$$

Atualizando o peso w_7 $\qquad w_7 = 0{,}5 \qquad \eta = 0{,}5$

$$w_7(t+1) = w_7 - \eta \frac{\partial E_T}{\partial w_7} = 0{,}5 - 0{,}5 \cdot (-0{,}0226025378)$$

$$= 0{,}5113012689$$