Currently, our update function is:

\[ a_i \leftarrow a_i - \alpha \frac{\partial L}{\partial a_i} \]

1. How would this update function change in order to incorporate weight decay.

\[
\text{Solution:} \\
\quad a_i \leftarrow (1 - \lambda \alpha) a_i - \alpha \frac{\partial L}{\partial a_i}
\]

2. Why would one use weight decay?

\[
\text{Solution:} \\
\quad \text{In order to do regularization. Weight decay provides a downward pressure on the parameters. This would reduce overfitting and increase generalization.}
\]