NOTE: you do not need a calculator for this quiz.
Assume you have a binary classification problem where \( f, \hat{y}, \) and loss function \( L \) are defined as:

\[
\begin{align*}
    f(x) &= a_0 + a_1 x + a_2 x^2 \\
    \hat{y} &= \sigma(f(x)) \\
    L(x, y) &= -(y \ln \hat{y} + (1 - y) \ln(1 - \hat{y}))
\end{align*}
\]

Let the learning rate, \( \lambda \), be .001, and the coefficients be:

\[
\begin{align*}
    a_0 &= -2 \\
    a_1 &= 1 \\
    a_2 &= 1
\end{align*}
\]

Assume one iteration of gradient descent with the training item:

\[
\begin{align*}
    x &= -2 \\
    y &= 1
\end{align*}
\]

Assume the simple optimization method: \( a_i = a_i - \lambda \frac{\partial L}{\partial a_i} \).

1. To reduce the loss, should each coefficient decrease (become more negative) or increase (become more positive) (circle one):

   1. \( a_0 \) decrease or increase?

   Solution: increase. Increasing \( a_0 \) increases \( f \) which increases sigmoid which increases \( \hat{y} \).

   2. \( a_1 \) decrease or increase?

   Solution: decrease. Decreasing \( a_1 \) increases \( f \) (since \( x \) is negative) which increases sigmoid which increases \( \hat{y} \).

   3. \( a_2 \) decrease or increase?

   Solution: increase. Increasing \( a_0 \) increases \( f \) which increases sigmoid which increases \( \hat{y} \).

2. What is the new value of \( a_1 \)?
Solution:

\[ f(x) = -2 + 1 \times -2 + 1 \times (-2)^2 = -2 - 2 + 4 = 0 \]
\[ \hat{y} = \sigma(f(x)) = \sigma(0) = 0.5 \]
\[ \frac{\partial L}{\partial f} = (\hat{y} - y) = (0.5 - 1) = -0.5 \]
\[ \frac{\partial f}{\partial a_1} = x = -2 \]
\[ a_1 = a_1 - \lambda \frac{\partial L}{\partial a_1} \]
\[ = a_1 - \lambda \frac{\partial L}{\partial f} \frac{\partial f}{\partial a_1} \]
\[ = 1 - .001 \times -0.5 \times -2 \]
\[ = 1 - .001 = 0.999 \]