On page 59 of Deep Learning with Python (2nd edition), it is said that grad(loss_val, x2) = 1. This note answers the question, "Why?"

For the example, y_true = 4 (actual value is 4) and x2 = 7 (predicted value is 7); so the value of the absolute error loss function is |4 – 7| = |-3| = 3.

We can view absolute error as a composition of 3 functions …

$$f(\hat{y}) = y - \hat{y}$$
$$g\big(f(\hat{y})\big) = (y - \hat{y})^2$$
$$h\Big(g\big(f(\hat{y})\big)\Big) = |y - \hat{y}| = \sqrt{(y - \hat{y})^2} = ((y - \hat{y})^2)^{\left(\frac{1}{2}\right)}$$

Using the chain rule with $h\Big(g\big(f(\hat{y})\big)\Big)$, where y_true is replaced with y and x2 is replaced by $\hat{y}$ …

$$\frac{\partial |y - \hat{y}|}{\partial \hat{y}}$$
$$= \frac{\partial ((y - \hat{y})^2)^{\left(\frac{1}{2}\right)}}{\partial \hat{y}}$$
$$= \frac{1}{2}((y - \hat{y})^2)^{\left(-\frac{1}{2}\right)} * \frac{\partial (y - \hat{y})^2}{\partial \hat{y}}$$
$$= \frac{1}{2}((y - \hat{y})^2)^{\left(-\frac{1}{2}\right)} * 2(y - \hat{y}) * \left(\frac{\partial y}{\partial \hat{y}} - \frac{\partial \hat{y}}{\partial \hat{y}}\right)$$
$$= \frac{1}{2}((y - \hat{y})^2)^{\left(-\frac{1}{2}\right)} * 2(y - \hat{y}) * (0 - 1)$$
$$= \frac{1}{2}((y - \hat{y})^2)^{\left(-\frac{1}{2}\right)} * 2(y - \hat{y}) * -1$$
$$= \frac{1}{2} * 2 * -1 * (y - \hat{y}) * ((y - \hat{y})^2)^{\left(-\frac{1}{2}\right)}$$
$$= -\frac{(y - \hat{y})}{((y - \hat{y})^2)^{\left(\frac{1}{2}\right)}}$$
$$= -\frac{y - \hat{y}}{|y - \hat{y}|}$$

For the general case …

$$\frac{\partial |y - \hat{y}|}{\partial \hat{y}} = \begin{cases} -1 & if\ y > \hat{y} \\ 0 & if\ y = \hat{y} \\ 1 & if\ y < \hat{y} \end{cases}$$