CS444 Gradient Descent Exercises

1. Calculate the partial derivatives of $f(x, y) = \sqrt{x^2 + y^2}$. Show your work.

2. Recall the least-squares error function for linear regression:

$$Error(\boldsymbol{w}) = \frac{1}{2}(y - \boldsymbol{w}^T \boldsymbol{x})^2$$

(This is the error associated with a single training sample with input $\boldsymbol{x}$ and target value y.)

This objective function encodes a belief that bigger errors are *much* worse worse smaller errors: in particular, that the penalty for making a mistake should grow with the square of the magnitude of the mistake. That seems reasonable[1], but it isn't the only possible error function. One problem with using a squared error function is that outliers can have a big impact on the result.

An alternative that is more robust to outliers is the absolute error (or L1 error):

$$\begin{aligned} Error(\boldsymbol{w}) =& |y - \boldsymbol{w}^T \boldsymbol{x}| \\ & |y - (w_0 x_0 + ... + w_n x_n)| \end{aligned}$$

Your goal in this exercise is to develop a gradient-descent learning rule for this new objective function. (It will be helpful to know that $\frac{d}{dx}|u| = \frac{u}{|u|} \times \frac{d}{dx}u$)

Your final rule should have the form:

$$w_i \leftarrow w_i - \eta???$$

[1] In fact, there are good statistical reasons for using this error function, particularly if the noise in the data is normally distributed.