Why does standardizing the original dataset help with gradient descent learning?