One of the most basic lower bounds in machine learning is that in nearly any nontrivial setting, it takes *at least* $1/\epsilon$ samples to learn to error ϵ (and more, if the classifier being learned is complex). However, suppose that data points are agents who have the ability to improve by a small amount if doing so will allow them to receive a desired positive classification. In that case, we may actually be able to achieve *zero* error by just being "close enough".

For example, imagine a hiring test used to measure an agent's skill at some job such that for some threshold θ, agents who score above θ will be successful and those who score below θ will not (i.e., learning a threshold on the line). Suppose also that by putting in effort, agents can improve their skill level by some small amount r. In that case, if we learn an approximation $\widehat{\theta}$ of θ such that

$$\theta \leq \widehat{\theta} \leq \theta + r$$

and use it for hiring, we can actually achieve error zero, in the sense that (a) any agent classified as positive is truly qualified, and (b) any agent who truly is qualified can be classified as positive by putting in effort. Thus, the ability for agents to improve has the potential to allow for a goal one could not hope to achieve in standard models, namely zero error.

In this paper, we explore this phenomenon more broadly, giving general results and examining under what conditions the ability of agents to improve can allow for a reduction in the sample complexity of learning, or alternatively, can make learning harder. We also examine both theoretically and empirically what kinds of improvement-aware algorithms can take into account agents who have the ability to improve to a limited extent when it is in their interest to do so.