# Why using regret in online learning

## 1. Worst-case cost and worst-case regret cost

Consider the function $g$ defined as $g(x) \triangleq \max_y f(x, y)$. The optimization problem defined as

$$\underset{x}{\text{minimize}} \quad g(x) \equiv \underset{x}{\text{minimize}} \quad \max_y f(x, y) \tag{1}$$

can be viewed as a worst-case optimization problem, and $g(x)$ is called *worst-case cost*. For example, $f(x, y)$ represents the time needed to drive a car from one place to another place, $x$ represents the driver's policy (speed up or not, the choice of roads), and $y$ represents it is raining or not. Then, $\max_y f(x, y) = f(x, \text{rain})$. Thus, $g(x)$ represents the time needed to drive a car from one place to another place in the worst case, i.e., raining days.

Now, consider $g'(x) \triangleq \max_y \left( f(x, y) - \min_{x'} f(x', y) \right)$. We call $g'(x)$ the *worst-case regret cost*. Unlike worst-case cost, the regret cost mesures the maximum achivement **it can be improved** in the past days. For example, if a driver cannot drive in the rainy days due to some reason. Then there is no difference between $f(x, \text{rain})$ and $f(x', \text{rain})$ for any $x$ and $x'$, which makes $g'(x) = 0$. No cost at all! Because nothing can be improved in the worst-case.

## 2. An example that minimize regret cost leads to a better policy

Consider the function $f(x, y)$ as follows:

| | | | |
|---|---|---|---|
| | | $x$ | $y$ |
| | | 0 | 1 |
| $x$ | 0 | 100 | 0 |
| $y$ | 1 | 100 | 99 |

The worst-case cost $g(x)$ in this example is $\max_y f(x, y) = f(x, 0) = 100$ for any $x$. Thus,

$$\underset{x}{\operatorname{argmin}}\, g(x) = 0 \vee 1. \tag{2}$$

But for the worst-case regret cost $g'(0) = 0$ and $g'(1) = 99$. Thus

$$\underset{x}{\operatorname{argmin}}\, g'(x) = 0. \tag{3}$$

## 3. Why using regret in online learning?

The key assumption of the tradition offline machine learning is that the data collected are independent and identically distributed from an unknown distribution [2]. This assumption can be easily violated in the field of online learning [1]. For example, consider again a driver that dose not drive during the rainy day. The data (time it spends from one place to another) point within 1 hour will be collected with probability 0. However, the data point within 1 hour will be collected with probability larger than 0. That means, the data collected in different dates follow different distributions. This makes a key difference between the traditional offline learning with online learning.

# BIBLIOGRAPHY


[1] Shai Shalev-Shwartz. Online Learning and Online Convex Optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.

[2] V.N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, sep 1999.