loss

Authors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Synonyms

cost

Abstract

The loss incurred by a hypothesis on a single data point is a scalar that quantifies the error of the hypothesis when evaluated on that data point. The ultimate goal of machine learning (ML) methods is to train a model such that the learned hypothesis delivers predictions with minimum loss. A key characteristic of ML applications and methods is the accessibility of loss values. In supervised learning, the loss can be evaluated for every hypothesis. In reinforcement learning (RL), the loss is observed only for the action actually taken.

Definition

In temperature forecasting, the quality of a predicted temperature is naturally measured by how far it deviates from the temperature actually measured the next day. The loss formalizes this idea. The loss incurred by a hypothesis $h \in \mathcal{H}$ on a single data point is a scalar that quantifies the error of h when evaluated on that data point (Shalev-Shwartz and Ben-David, 2014, Sect. 2.1). A smaller loss value indicates a smaller error. In supervised learning, this error is the discrepancy between the prediction $h(\mathbf{x})$ delivered for the features $\mathbf{x}$ and the label y of that data point. Some losses (e.g., the squared error loss) take a continuum of values, while others (e.g., the 0/1 loss) are binary, indicating only whether the prediction matches the label. For the temperature forecast, the loss on a given day can be the squared difference between the predicted temperature (the prediction) and the measured one (the label); see Fig. 1.

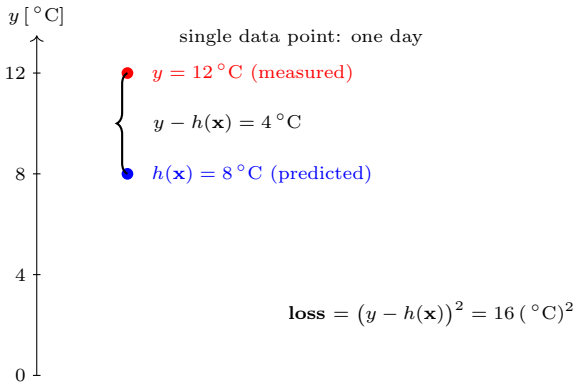

Figure 1: Loss on a single data point representing one day. The true label is the measured maximum daytime temperature, $y = 12\,^\circ\mathrm{C}$; the hypothesis predicts $h(\mathbf{x}) = 8\,^\circ\mathrm{C}$. Under the squared error loss, the loss incurred on this data point is $\big(y - h(\mathbf{x})\big)^2 = 16\,(\,^\circ\mathrm{C})^2$

Loss values are typically nonnegative: many losses are built from a norm or metric and are therefore bounded below by zero. Nonnegativity is not required, however: the logarithmic loss (negative log-likelihood) underlying maximum-likelihood methods can take negative values, since a probability density can exceed one. The loss is also not specific to supervised learning. In unsupervised learning there is no label: a clustering loss measures how far a data point lies from its assigned cluster centroid, and the loss of an autoencoder is the reconstruction error between a data point and its reconstruction.

Machine learning (ML) methods use observed loss values to construct an objective function for model training. The average loss across a training set of data points is the empirical risk that empirical risk minimization (ERM) minimizes.

ML applications can be categorized by how the loss values are accessible. In supervised learning, the loss can, in principle, be evaluated for every hypothesis in the hypothesis space, since the true label of each training data point is observed. In reinforcement learning (RL), the situation differs (Sutton and Barto, 2018). At each time step, the ML method (implemented by an agent) selects an action via the currently used hypothesis. The agent then observes the loss incurred by the action actually taken. The loss that would have been incurred by the other actions is not revealed. This limited accessibility of loss values is known as bandit (or partial) feedback (Hazan, 2022, Sect. 6.1). For example, a route-planning system learns the actual travel time only for the route the driver took, not for any alternative route. Similarly, a self-driving car selects a steering direction, and the incurred loss can be measured from onboard sensors such as collision detectors and lane-departure warnings, but only for the trajectory actually driven.

Cross References

loss function, logarithmic loss, data point, hypothesis, empirical risk, ERM, norm, metric, RL, supervised learning, unsupervised learning

References

1. Hazan (2022). Introduction to Online Convex Optimization. 2nd ed., MIT Press, Cambridge, MA, USA.

2. Shalev-Shwartz and Ben-David (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge Univ. Press, New York, NY, USA.

3. Sutton and Barto (2018). Reinforcement Learning: An Introduction. 2nd ed., MIT Press, Cambridge, MA, USA.