Title: Structural risk minimization

URL: https://en.wikipedia.org/wiki/Structural_risk_minimization

PageID: 10704974

Categories: Category:Machine learning, Category:Machine learning stubs

Source: Wikipedia (CC BY-SA 4.0).

Structural risk minimization (SRM) is an inductive principle of use in machine learning . Commonly in machine learning, a generalized model must be selected from a finite data set, with the consequent problem of overfitting – the model becoming too strongly tailored to the particularities of the training set and generalizing poorly to new data. The SRM principle addresses this problem by balancing the model's complexity against its success at fitting the training data. This principle was first set out in a 1974 book [1] by Vladimir Vapnik and Alexey Chervonenkis and uses the VC dimension .

In practical terms, Structural Risk Minimization is implemented by minimizing E t r a i n + β H (W) {\displaystyle E_{train}+\beta H(W)} , where E t r a i n {\displaystyle E_{train}} is the train error, the function H (W) {\displaystyle H(W)} is called a regularization function, and β {\displaystyle \beta } is a constant. H (W) {\displaystyle H(W)} is chosen such that it takes large values on parameters W {\displaystyle W} that belong to high-capacity subsets of the parameter space. Minimizing H (W) {\displaystyle H(W)} in effect limits the capacity of the accessible subsets of the parameter space, thereby controlling the trade-off between minimizing the training error and minimizing the expected gap between the training error and test error. [2]

The SRM problem can be formulated in terms of data. Given n data points consisting of data x and labels y, the objective J (θ) {\displaystyle J(\theta)} is often expressed in the following manner:

J (θ) = 1 2 n ∑ i = 1 n (h θ (x i) – y i) 2 + λ 2 ∑ j = 1 d θ j 2 {\displaystyle J(\theta)={\frac {1}{2n}}\sum _{i=1}^{n}(h_{\theta }(x^{i})-y^{i})^{2}+{\frac {\lambda }{2}}\sum _{j=1}^{d}\theta _{j}^{2}}

The first term is the mean squared error (MSE) term between the value of the learned model, h θ {\displaystyle h_{\theta }} , and the given labels y {\displaystyle y} . This term is the training error, E t r a i n {\displaystyle E_{train}} , that was discussed earlier. The second term, places a prior over the weights, to favor sparsity and penalize larger weights. The trade-off coefficient, λ {\displaystyle \lambda } , is a hyperparameter that places more or less importance on the regularization term. Larger λ {\displaystyle \lambda } encourages sparser weights at the expense of a more optimal MSE, and smaller λ {\displaystyle \lambda } relaxes regularization allowing the model to fit to data. Note that as λ → ∞ {\displaystyle \lambda \to \infty } the weights become zero, and as λ → 0 {\displaystyle \lambda \to 0} , the model typically suffers from overfitting.

See also

Vapnik–Chervonenkis theory

Support vector machines

Model selection

Occam Learning

Empirical risk minimization

Ridge regression

Regularization (mathematics)

References

External links

Structural risk minimization at the support vector machines website.

v

t

e