

What is Empirical Risk Minimization

AUGUST 19, 2017AUGUST 18, 2017 ~ PRATEEK JOSHI



Even though it has an ornate name, the underlying concept is actually quite simple and intuitive. The concept of Empirical Risk Minimization becomes relevant in the world of supervised learning. The actual goal of supervised learning is to find a model that solves a problem as opposed to finding a model that best fits the given dataset. Since we don't have every single data point that represents each class completely, we just use the next best thing available, which is a dataset that's representative of the classes. We can think of the process of supervised learning as choosing a function that achieves a given goal. We have to choose this function from a set of potential functions. Now how can we measure the effectiveness of this chosen function given that we

don't know what the actual distribution looks like? Bear in mind that all the potential functions can achieve the given goal. How do we find the function that's the best representative of the true solution?

Understanding the concept of risk

To understand it, we need to talk a bit about the idea of a loss function. Given a set of inputs and outputs, this loss function measures the difference between the predicted output and the true output. But this is applicable only to the given set of inputs and outputs. We want to know what the loss is over all the possibilities. This is where "true risk" comes into picture.

True risk computes the average loss over all the possibilities. But the problem in the real world is that we don't know what "all the possibilities" would look like. In mathematical terms, we say that we don't know the true distribution over all the inputs and outputs. If we did, then we wouldn't need machine learning in the first place.

Give me an example

For example, let's say you want to build a model that can differentiate between a male and a female based on certain features. If we select 100 random people where men are really short and women are really tall, then the model might incorrectly assume that height is the differentiating feature. To build a truly accurate model, we need to gather all the men and women in the world to extract the differentiating features. Unfortunately, that's not possible! So we select a small number of people and hope that this sample is representative of the whole population.

What exactly is empirical risk minimization?

We assume that our samples come from this distribution and use our dataset as an approximation. If you compute the loss using the data points in our dataset, it's called empirical risk. It's "empirical" and not "true" because we are using a dataset that's a subset of the whole population.

When we build our learning model, we need to pick the function that minimizes the empirical risk i.e. the delta between the predicted output and the actual output for the data points in our dataset. This process of finding this function is called empirical risk minimization. Ideally, we would like to minimize the true risk. But we don't have the information that allows us to achieve that, so our hope is that this empirical risk will almost be the same as the true empirical risk. Hence by minimizing it, we aim to minimize the true risk.

What does it depend on?

The size of the dataset has a big impact on empirical risk minimization. If we get more data, the empirical risk will approach the true risk. The complexity of the underlying distribution affects how well we can approximate it. If it's too complex, we would need more data to get a good approximation. We should also be careful about the family of functions we consider. If the size is too large, then the approximation error will be very high in certain situations. The behavior of the loss function itself can impact it. If we are not careful in choosing this function, then we might end up with very high loss values. L2 regularization is a very good example of empirical risk minimization.

POSTED IN [MACHINE LEARNING](#)

[DISTRIBUTION](#) [PROBABILITY](#) [SUPERVISED LEARNING](#)



Published by Prateek Joshi

Founder of Plutoshift. Author of 13 books on Machine Learning. Featured on Forbes, NBC, Bloomberg, CNBC, TechCrunch, Silicon Valley Business Journal, and more. Speaker at conferences such as TEDx, Global Big Data Conference, Machine Learning Developers Conference, and Sensors Expo. You can learn more about me here: www.prateekj.com [View all posts by Prateek Joshi](#)

3 thoughts on “What is Empirical Risk Minimization”

1. **[Gaurav Bhatia](#)**

SAYS:

[SEPTEMBER 4, 2017 AT 12:59 AM](#)

Thank you for posting this helpful blog on empirical risk minimization. I gained a lot of info from this blog and I can't wait to try this on my own. Looking forward to read your next post.

[Reply](#)

2. **[amira05](#)**

SAYS:

[MARCH 28, 2018 AT 11:08 AM](#)

thank u for this wonderful explanation very good teacher ♥

[Reply](#)

3. **[Badarinath C](#)**

SAYS:

[OCTOBER 5, 2018 AT 12:43 PM](#)

Thaks a lot to explaining difference between empirical risk and true risk

[Reply](#)

BLOG AT WORDPRESS.COM.