Task：知乎

#Before we get started to introduce MAML, I’d like to give you guys a first glance of the essence of Meta Learning. We all know what’s the goal in the ordinary machine learning, which is to find an appropriate function to fit the dataset in order to get the best accuracy. However, for meta learning, intuitively I can describe it as to find a function F, which can output different parameter matrices, regarded as another function f, to best fit different tasks provided by the user. In other words, machine learning focuses on learning a model to predict something, and meta learning focuses on how to learn a predictive model faster and better, also can be understood by learn2learn.

Machine learning 在一个distribution上训练与预测，meta learning在不同的distribution上的task下训练，e.g cat distri, dog distribution. Sin函数平移例子。

Meta parameters can do fine-tuning

Let's take an example in our daily lives. When we teach children to read English, we can directly ask them to imitate the pronunciation of apple, banana. But soon they'll come across new words, like strawberry, and they'll have to listen to you again to hear the words correctly. Let's do it differently, this time instead of teaching the pronunciation of individual words, let's teach the pronunciation of phonetic symbols. From then on, when children meet new words, they can pronounce the word correctly as long as they follow the phonetic symbols. The process of learning phonetic symbols is a meta-learning process.

Since my teammates have introduced tasks in meta-learning, I’ll explain the MAML algorithm quickly. MAML is a typical type of algorithm in the meta learning field. Its goal is to train a best a set of parameters to obtain a fast adaption in different test scenarios, as showing in the ppt slides. We can learn good parameters, which can quickly converge to the global minimum in different tasks. 更多的explaination在这个图。For the concrete algorithm, I will briefly explain it as doing two different stochastic gradient descent processes. First inner GD is done on each task in each epoch and get an inner updated parameter of a copied model, the second GD is done to update the original model based on the updated parameter to get parameter . Repeat the process until convergence. Actually, we update the parameters of our model only on the second GD. By the illustrations above, we can obtain the result that only a few new tasks can make MAML converges rapidly.

看图说话。Meta-train：最有潜力的方向 and meta-test：用support set fine tuning，用updated θ做最终的测试。

Sin函数different domains。拉回widar

However, since our goal is to clearly classify images provided in Widar, it has different domains. From the theoretical perspective, we only have an upper bound to guarantee that the MAML algorithm will converge, but the bound may not be as tight as possible. Then, in order to do a more precise classification on images from different domains, we introduced MAML-DG. In the meta-training stage, the main difference is that we update the parameters by sampling tasks from two different domains, which empirically guarantees that the model will converge on different domains. In other words, we use the loss function to guarantee the convergence, rather than the upper bound. We first generate GD twice in the first domain, and generate another GD in another domain. From the new algorithm, we can let the model learn the distribution shift from the dataset of different domains. Also, we introduced a meta-testing stage in the process. When we have a new task generated from a new domain, we do the fine-tuning using the support set and then use the query set to test the model. Eventually we find that the model can precisely find the distribution shift among the new task and the previous tasks in the training set, which implies the fast convergence will be guaranteed also in the domain generalization field.