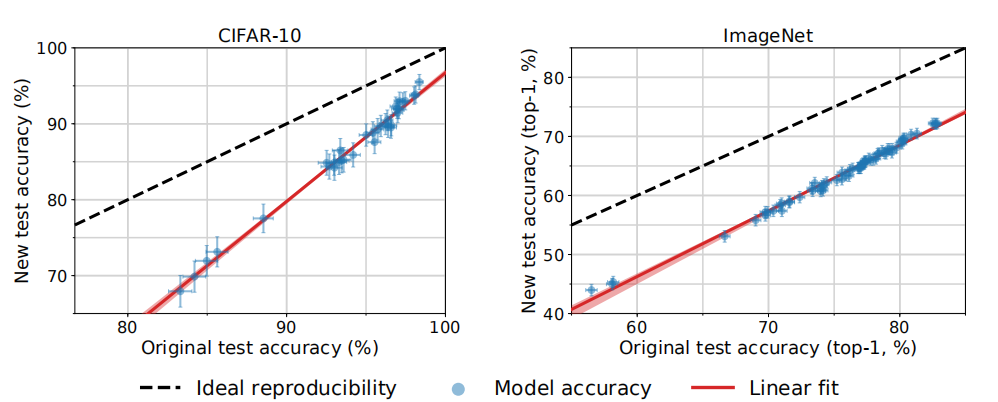
This doc helps record my understanding of the paper.

**1 Introduction**

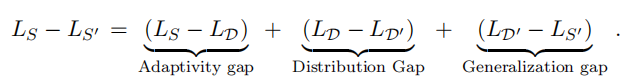


The black dash line shows the ideal situation when we apply the same model on old dataset (where the model is trained) and new dataset, they should have same performance. And the red line is the real situation. We can witness the great drop in both CIFAR-10 and ImageNet. One thing interesting is that, **the model performance would keep their order.**

Personally, I think both CIFAR-10 and ImageNet are huge enough to represent the real pixel distribution of an item. So the model which performs better on these two dataset can be considered as generalized model. That’s why when models are applied on another data, better models are still better.

#### 2 Potential Causes of Accuracy Drops

In this section, there is a decomposition of the formula.



In this formula, we we have **D the original data distribution**, **S original data samples**. That’s the basic idea of machine learning: we can hardly get the data which can represent or match the total distribution, so we assume our sample **S** is follow **D** distribution.

The first part of decomposition describe this gap(**Adaptivity gap**). And the third element has also the same meaning, the only difference is the calculation sign. There is no doubt that difference can also be seen between two dataset **D** and **D’**,it will definitely affects the model performance**.** In this paper, the author mark them as Distribution Gap. They believe that this is the most significant reason that cause the performance drop.

**3 4 5**

In these part, they try to illustrate how they select new data and how they do experiment on them. They apply a parameter describing the difficulties of one image to be correctly classified. Meanwhile, for every model, there is a parameter to describe the ‘skill’ of one model,higher skill means higher accuracy.

They build a gaussian distribution associated with these two parameters to describe a data set. Then one can use statistics to compare two data set. Then author build a linear relationship to find out the underlining of accuracy drop.

For me , I think this paper it’s incomplete.

For example, the shape and function of mobile phones are dramatically different between today’s devices and ten years ago. But there is still some relationship between them so one may not consider a very old mobile phone as a telephone.

It’s the same when it comes to model and image data set. The shape of items may change a lot with time but they would still share some common features. Let’s say the features changes 20%, then it’s nature to think the accuracy would drop, and the drop value should be a function of this “20%”. In this paper, they observe this drop and fit the drop with a linear model.

But what if this linear relationship is a mixture function? It still remains a interesting mystery.