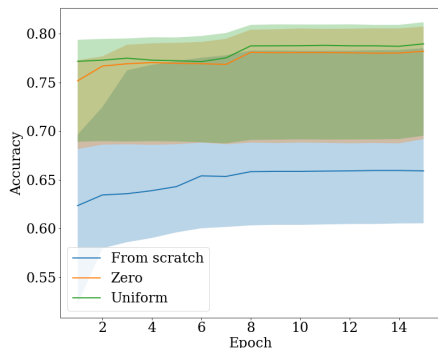


Anti-Distillation from a simple model to a complex one.

Goal: Adapting the model to more complex data.



Increasing complexity:

1. More classes
2. More features

Our solution: Weight initialization using the parameters of the pre-trained model.

$$\mathbf{w}_1 = \varphi(\mathbf{u}^*)$$

$$\mathbf{w}_1 = (\mathbf{u}^*, \mathbf{w}'_1), \quad \mathbf{w}'_1 = \mathbf{0}, \quad \mathbf{w}'_1 \sim \mathcal{U}\left[-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right]$$