PadhAI: Regularization

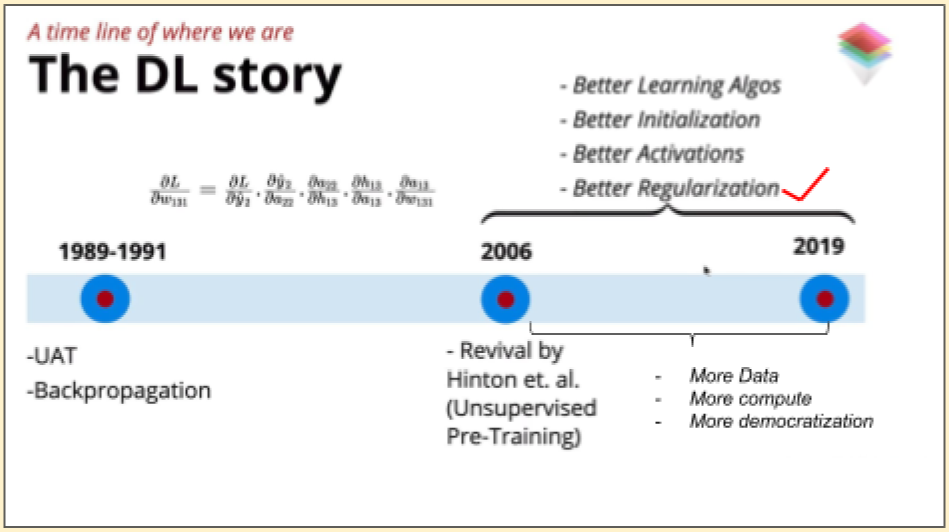
One Fourth Labs

Regularization Methods

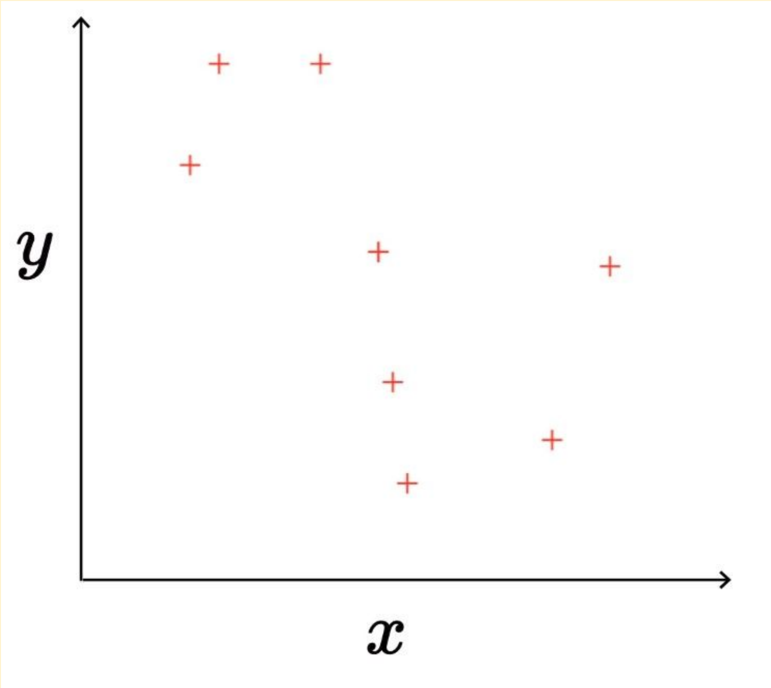
Simple vs complex models

The timeline of where we are

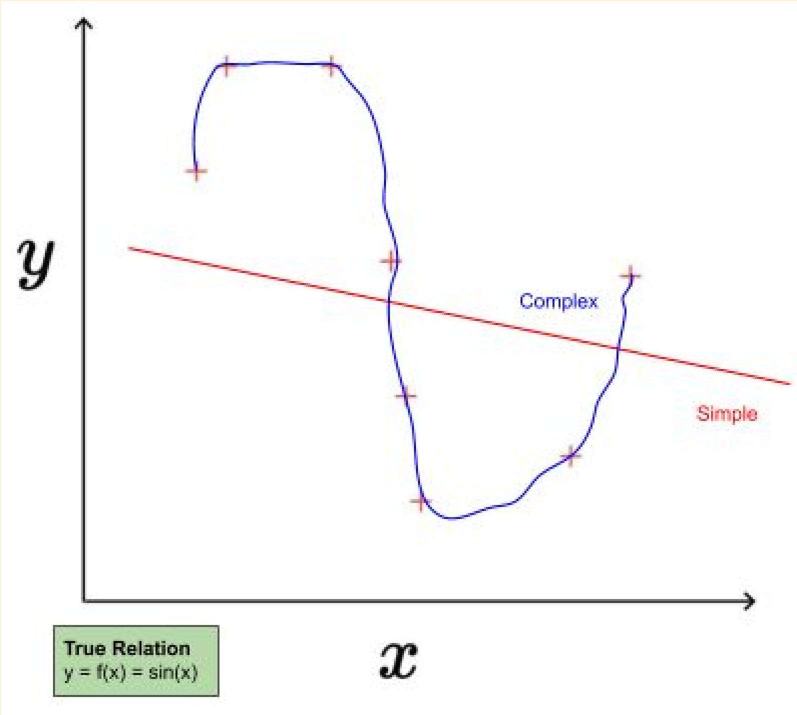
1. In this section, we will look at how better Regularization methods have accelerated the growth of DL over the last decade



2. Why do we need **Regularization**?
   a. To answer this question, we must look at a concept known as **Bias Variance trade-off**. The Bias that we're speaking about here is different from the bias parameter b that we have seen so far in Neural Networksμ
   b. Consider the following toy data visualisation



c. In the above figure, the true relation is **y = f(x), where f(x) = sin(x)**, however, in practice, that is not known to us. So we try to approximate models.



d. **Simple**(degree 1): $y = \hat{f}(x) = w_1 x + w_0$
   i. We assume that the relationship between y and x is a straight line of the form mx + c
   ii. This looks like a very naive assumption.
   iii. It is represented by the Red line in the figure
   iv. The best fitting Red line is plotted while trying to minimize the error/loss between the predicted points and the actual points
   v. This is a pretty bad model, where even the minimised loss is still far too high

e. **Complex**(degree 25): $y = \hat{f}(x) = \Sigma^{25}_{i=1} w_i x^i + w_0$
   i. This is a degree 25 polynomial, with 26 parameters (including w_0)
   ii. It is represented by the Blue curve in the figure
   iii. The Blue curve is plotted the same way, by minimising the error/loss between predicted and actual values
   iv. Here, there is zero error/loss, it is a perfect fit.

3. Now, how does this relate to Bias and Variance and how does it in turn lead to regularization.