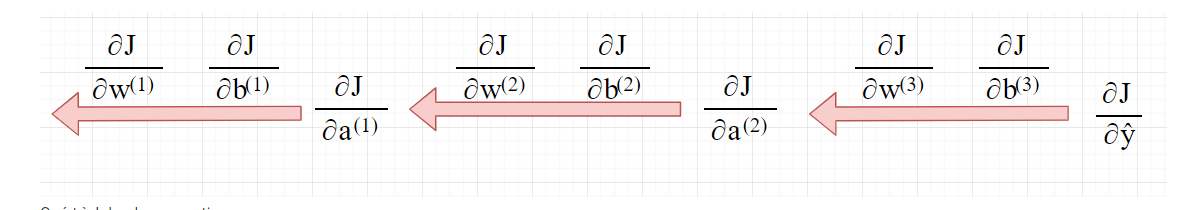
So far, we’ve only talked about linear models: linear regression and linear binary classifiers. We noted that there are functions that can’t be represented by linear models; for instance, linear regression can’t represent quadratic functions, and linear classifiers can’t represent XOR. We also saw one particular way around this issue: by defining features, or basis functions. E.g., linear regression can represent a cubic polynomial if we use the feature map ψ(x) = (1, x, x2 , x3 ).

Backpropagation is the core algorithm that makes deep learning models computationally easy. With modern NN networks, thanks to this algorithm, the optimization algorithm with gradient descent can be millions of times faster than the traditional implementation. Just imagine a model with back-propagation that takes a week to run can take up to 200,000 years to train with traditional methods!

Although backpropagation is used for deep learning, it is also a powerful computational tool for many other fields from weather forecasting to numerical stability analysis, only it is used with different names. In fact, it was rediscovered to be used in many different fields. But in general, regardless of the application, its name is "reverse-mode differentiation".

It's basically a technique for quickly calculating the derivative. And it is an essential trick that you need to prepare yourself not only in the field of deep learning but also for many other arithmetic computation problems.



We apply two algorithms for two kind of problem, regression(cost function) and classification(cross-entropy).