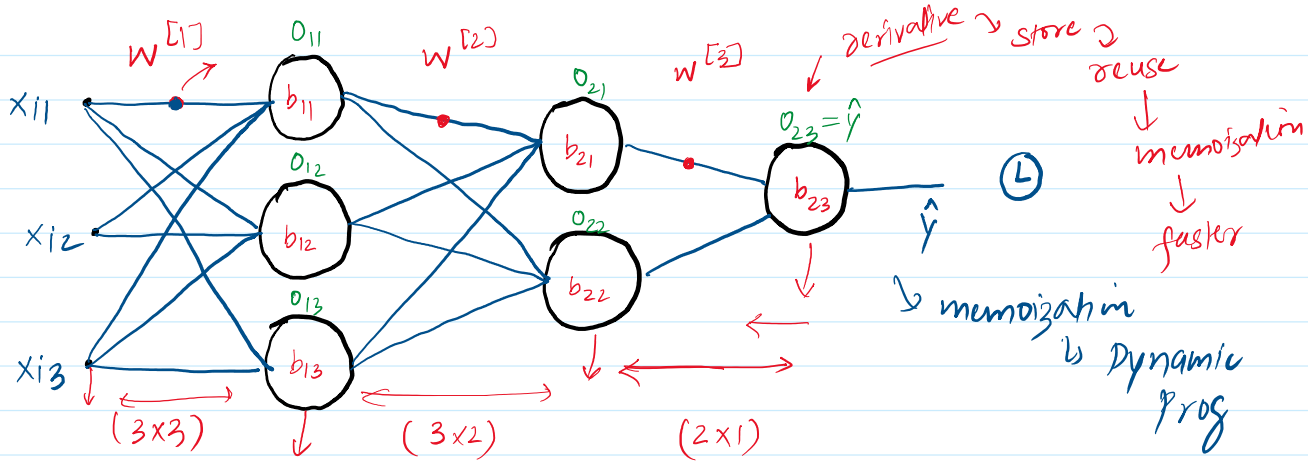
MLP Memoization

Thursday, April 7, 2022 7:45 AM

backprop update → w / b) deriv

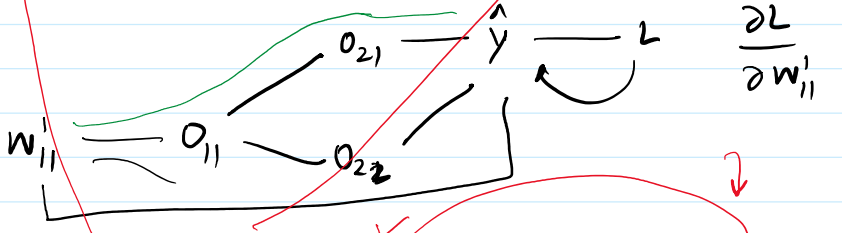


Chain rule

$L \to \hat{y} \to O_{21} \to O_{11} \to w^1_{11}$

$$\frac{\partial L}{\partial w^3_{11}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w^3_{11}} \qquad \frac{\partial L}{\partial w^2_{11}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial w^2_{11}}$$

$$x \to f(x), g(x) \to h(f(x), g(x)) \qquad \frac{\partial h}{\partial x} = \left[\frac{\partial h}{\partial f(x)} \times \frac{\partial f(x)}{\partial x}\right] + \left[\frac{\partial h}{\partial g(x)} \times \frac{\partial g(x)}{\partial x}\right]$$



$$\left[\frac{\partial L}{\partial w^1_{11}} = \frac{\partial L}{\partial \hat{y}}\left[\frac{\partial \hat{y}}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial w^1_{11}} + \frac{\partial \hat{y}}{\partial O_{22}} \times \frac{\partial O_{22}}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial w^1_{11}}\right]\right]$$

Backprop → Chain diff (rule) + Memorization

Keras TF — maths — (3)