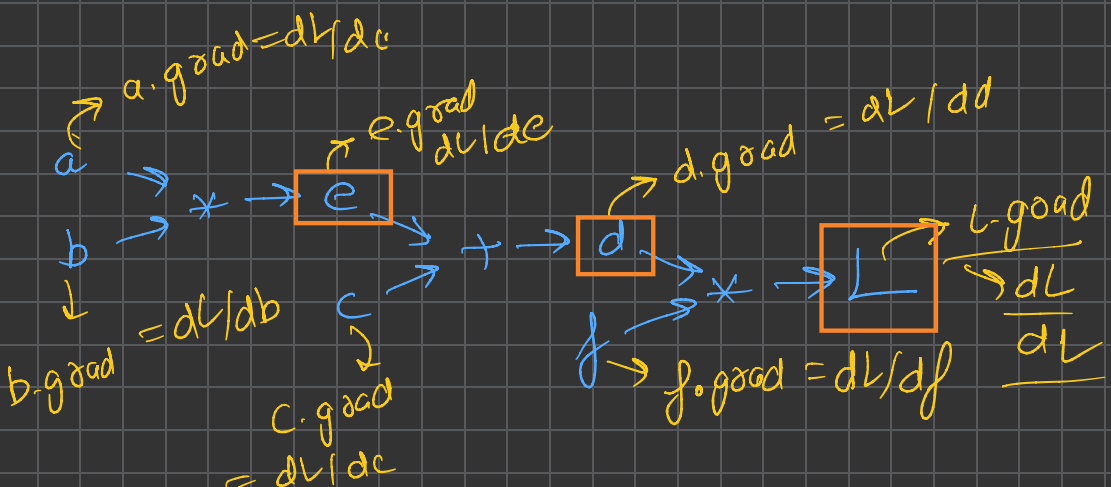


$e = a * b$, $d = c + e$, $L = d * f$.

Forward Pass

$e = a * b$

$d = c + e$

$L = d * f$

(will we be, updating intermediate nodes, like, e, d, L, as we have calc. grad for them)

Answer:- we calculated grad for all var. , a, b, c, d, e, f, L. , but we will only update, a, b, c, f. , but not e, d, L , bcz, e, d, L are intermed.

are intermediate steps, if we even ~~also~~ update c, d, L, it won't make any sense, we will waste ops. if $c = a * b$, & we have updated, a & b, then even if we update c, it won't matter, as forward pass, we will overwrite, the value of c;

if $a \Rightarrow a+h$, $b \Rightarrow b-h$, $c \Rightarrow c+h'$ → won't make sense

$c = (a+h)(b-h)$,

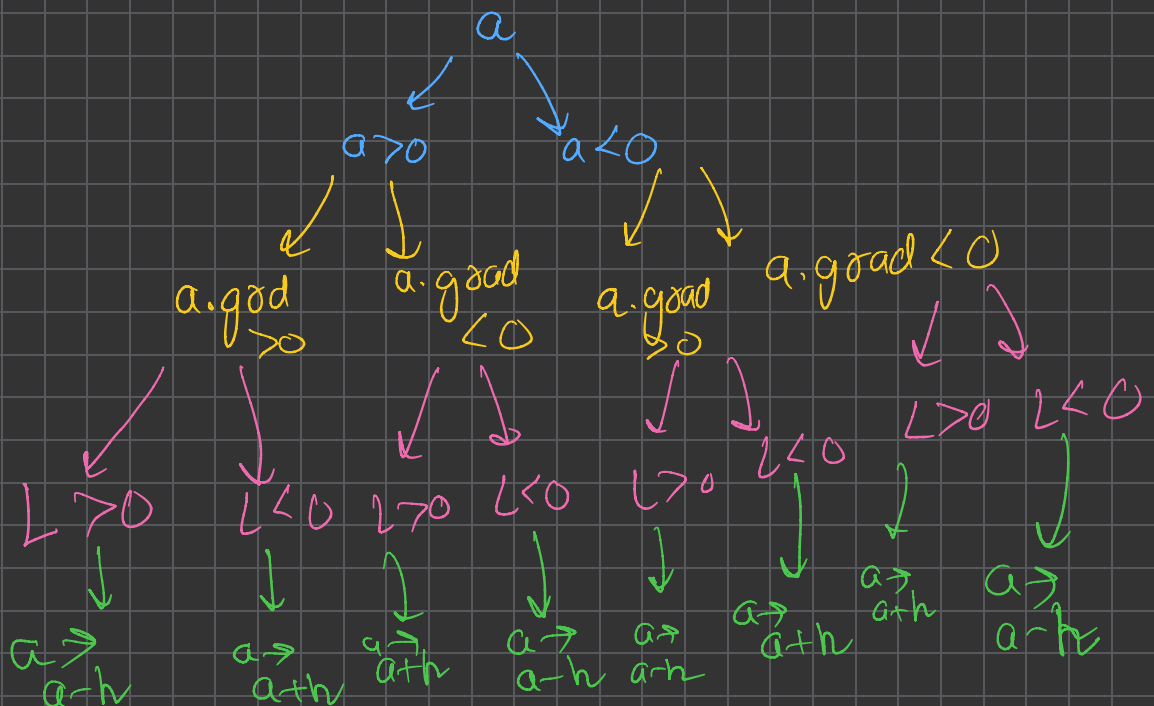
The value of $c+h'$, will be overwritten by $(a+h)(b-h)$

↳ conclusion:- we will update only trainable parameters, not the intermediate parameters

Eg (a, b, c, f)

General, Assume a variable 'a' is there

⟹



If ignoring, $a>0$, or $a<0$, (Assume, $L<0$)

if $a.grad > 0$, then, $\Rightarrow a = a+h$

$a.grad < 0$, then, $\Rightarrow$ ~~or~~ $a = a-h$

$a \mathrel{+}= \alpha\, a.grad$ (Assume $h = a.grad$)

$a = a + (\alpha.(a.grad)$ $(a.grad > 0, < 0)$

$a = a + (\alpha.(a.grad)$ $(a.grad > 0)$

$a = a - (\alpha.(a.grad)$ $(a.grad < 0)$

$\rightarrow \alpha = 0.00$ $(h = \alpha.(a.grad)$

if $L < 0$, conclusion, update parameters in dirⁿ of gradient.

if $L > 0$, conclusion, update parameter in opp. dirⁿ of grad.

if $L < 0$

$a.data \mathrel{+}= \alpha (a.grad)$

↳ learning rate

if $L > 0$

$a.data \mathrel{-}= \alpha (a.grad)$

↳ learning rate