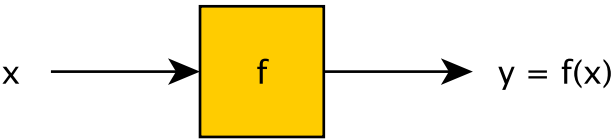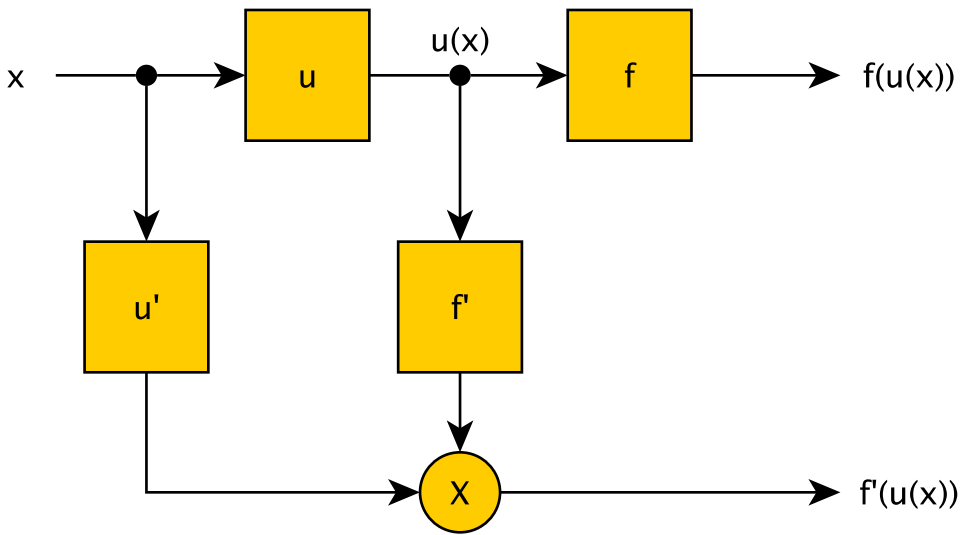**function f maps x to y**
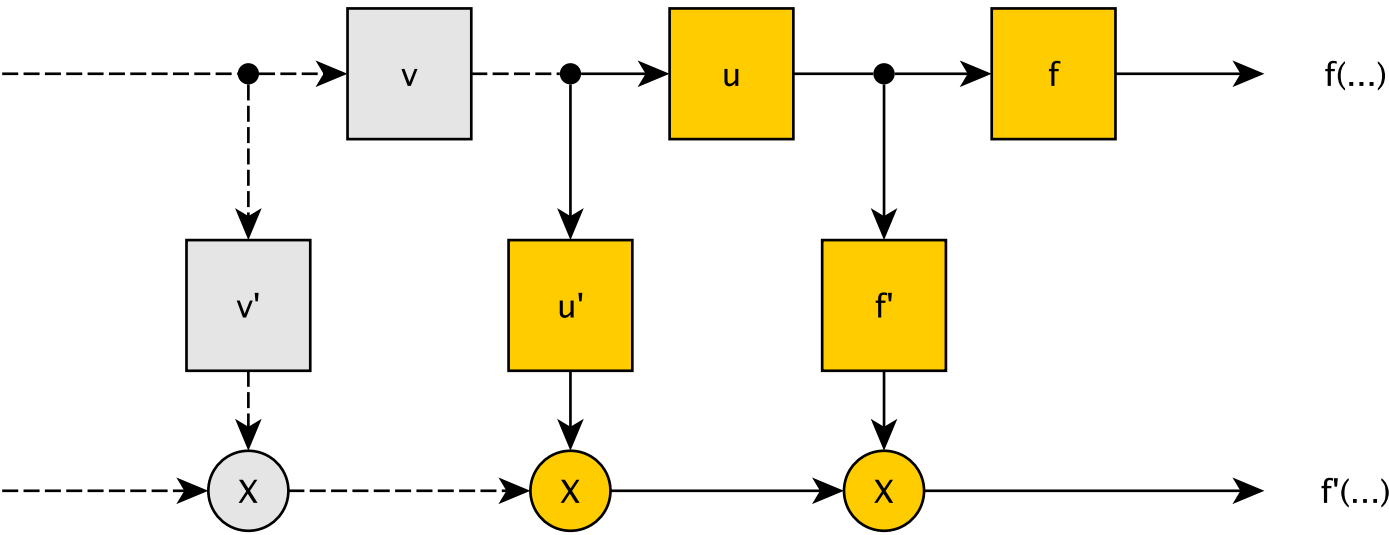
x → [f] → y = f(x)

Chain rule

$f(u(x)) \rightarrow f'(u(x)) = f'(u(x)) * u'(x)$

more generalized:
add a branch for each function composition

Remember the subtle differences:

– look at f(x) as a function of a weight, not of input data
– there is a difference in function composition:
   – f(u(x1,x2))
   – f(u(x1), x2)