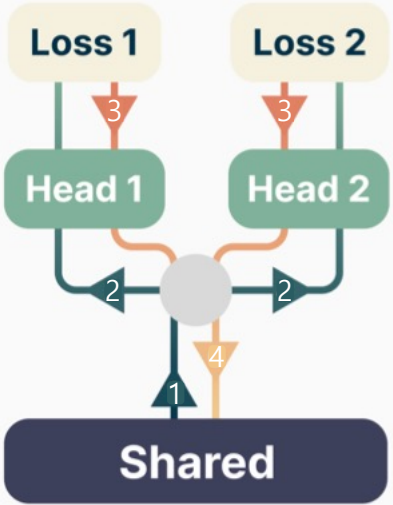
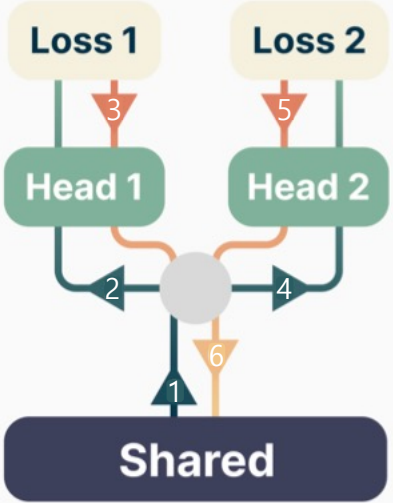


(a) Naïve




```
z = model.shared(x)
d = z.detach()
d.requires_grad = True

for i in range(n):
    p = model.heads[i](d)
    loss(p, y[i]).backward()

z.backward(gradient=d.grad)
```

(b) Memory Efficient