1. 前向传播与反向传播

之前用SVM损失函数与softmax分类器计算的都是得分以及概率，并得到相应的损失函数。这是前向传播的过程。

得到损失函数以后，我们要根据损失函数来优化我们的模型，即优化W参数，更新W参数，使得损失值越低越好。这是反向传播的过程。

1. 优化模型

那么，我们如何来优化W参数呢？这里，我们就要计算每个W参数对于最终的损失值的重要程度。若某个W参数越大，损失值也越大，则说明应该将W参数减小。若某个W参数越大，损失值也越小，则说明应将W值增大。

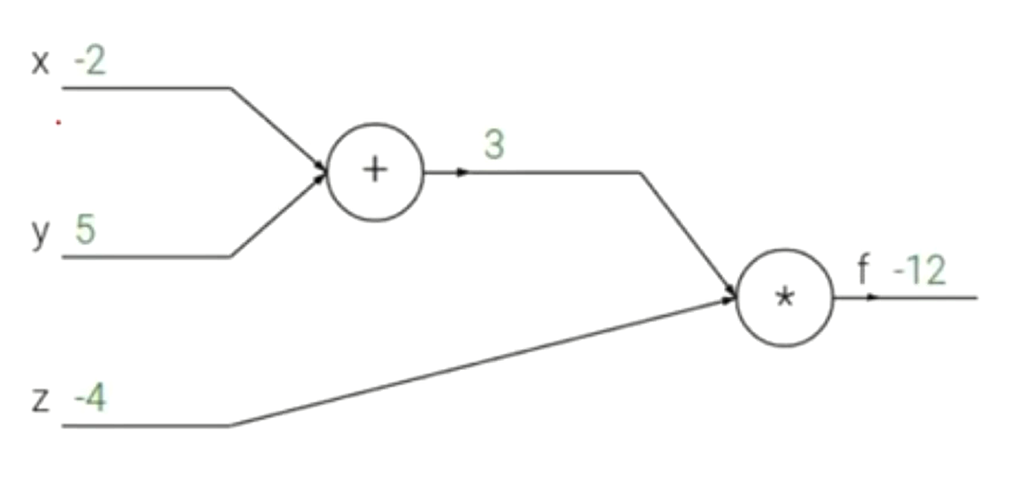
举例：

（一）损失值函数：

f(x, y, z) = (x + y)z

有三个样本，

X = -2, y = 5, z = -4

画出计算图：  


（二）计算反向传播

现在开始计算x,y,z分别对f值的贡献程度“

我们设q = x + y，可以先求出x和y分别对q的贡献程度，即

Dq / dx = 1

Dq / dy = 1

然后，计算q对于f的贡献程度：

F = (x+y)z = qz

Df / dq = z = -4

则根据链式法则，可以求出x，y分别对f的贡献程度：  
df / dx = df / dq \* dq / dx = -4 \* 1 = -4

即x的值增大一倍，f的值就要减小4倍

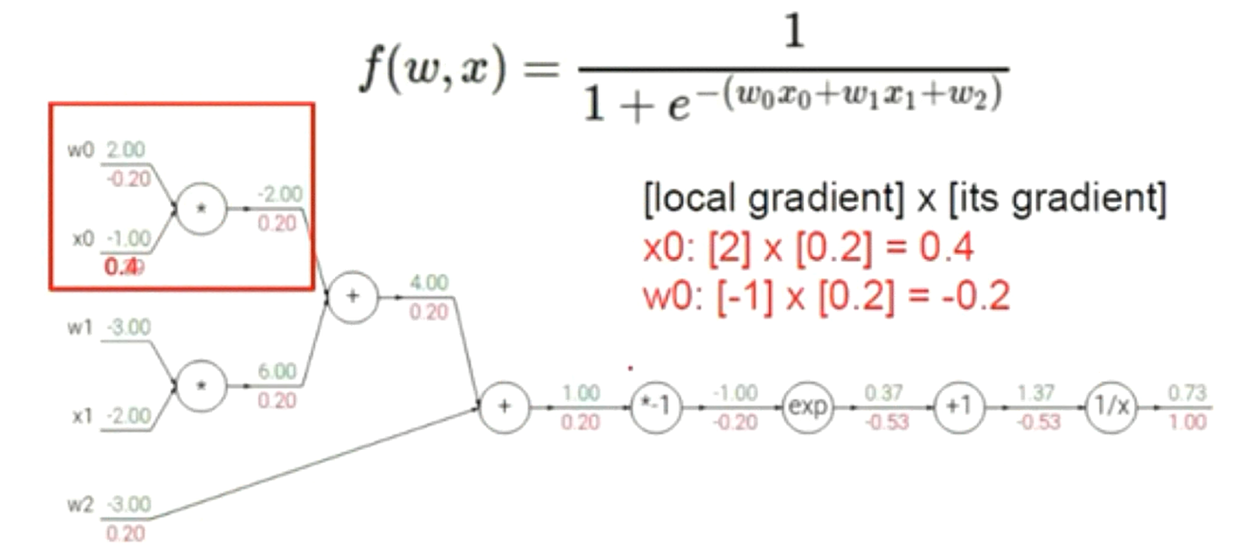
Df / dy = df/ dq \* dq / dy = -4 \* 1 = -4  
即y的值增大一倍，f的值就要减小4倍

计算z对f的贡献程度，即计算f对于z的偏导：

Df / dz = q = x+y = -2 + 5 = 3

即z的值增大一倍，f的值就要增大三倍

1. Sigmoid模型反向传播



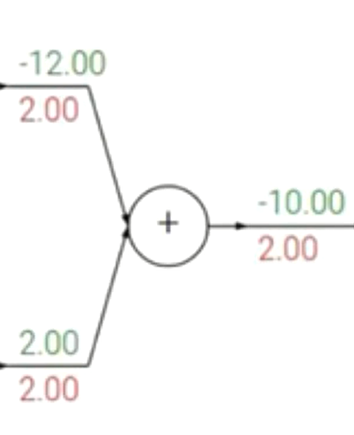
这是一个sigmoid函数的计算图，前向传播用绿色数字标识在箭头上面，反向传播用红色数字标识箭头下面。

我们从反向传播来看，结尾处的计算是 f = 1/x，则 df/ dx =-1/x^2^，且前向传播输入的x是1.37，则反向传播就要用输入的x来计算，df / dx = -1 / (1.37)^2^，得到的结果是 -0.53。

这是一个需要注意的地方，通过这样运算，反向传播就能够从损失值一直传回到各个参数，计算出各个参数对于损失值的影响程度。

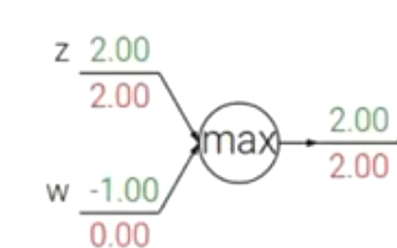
四 . 门单元：  
加法门单元：

均匀分配，各个参数的反向传播梯度是一样的。



Max门单元：

前向传播中Max门单元将最大值传播了过去，那么反向传播，梯度只会分配给最大值



乘法门单元：

反向传播的梯度分配是按照前向传播的参数互换后的比值来分配的，

