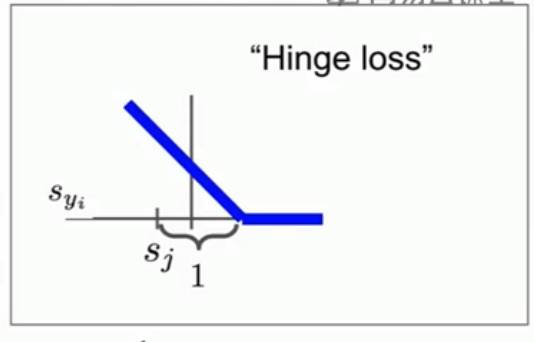
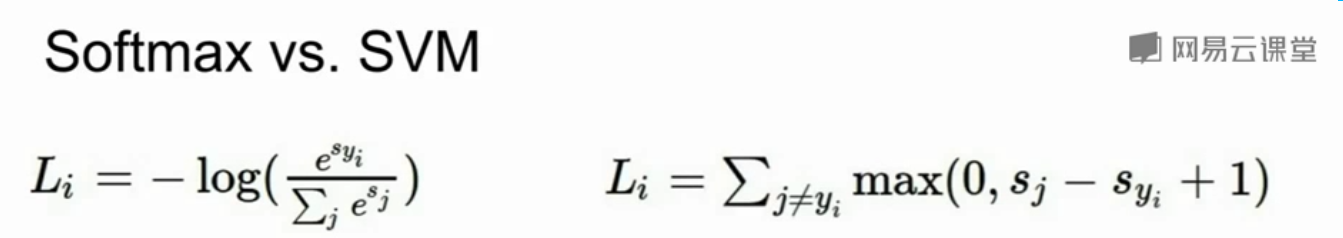
# 多分类SVM分类

loss值计算方式

* 当正确分类的分数比其他任意错误分类的分数高于一个阈值，则损失为零否则损失为错误分类的分数加一减去正确分类分数
* 将上述对于每个错误分类计算出的损失相加既可以得到整体的损失
* 计算整个训练集的loss求平均
* 该loss为合页损失函数



1. 这里的1被称为safety margin安全边界
2. scale是什么？
3. That this free parameter of one kind of washes out and is canceled with this scale like the overall setting of the scale in W
4. 对于这个loss在开始迭代的时候应该为C-1，因为初始的W值比较平均，导致计算出的score值较平均，loss为1，求和之后是C-1（C是所有的类别）
5. multinomial logistic regression（多项逻辑斯蒂回归损失函数），将score指数化使结果为正数，之后归一化



1. SVM的loss计算方法与softmax的区别，SVM的loss计算中当正确分类分数超过错误分类一定数值的时候loss为零，即不再向更好的方向优化。Softmax loss不会为零而是不断向更好的方向优化
2. 正则化，包括添加l1正则化（更加倾向于稀疏的矩阵），l2正则化（更加倾向于将整体的多项式的最高次数降低）

# 优化算法

为了寻找最好的w计算得到最小的loss，需要采取优化的算法，然而对于大量参数来说，求解最小值较为复杂，因此对于如何得到一个较好的w有一些策略对应算法

策略1：随机生成w计算loss

策略2：跟随斜率

1. 在一维函数f(x)中，求解斜率就是对f(x)求导
2. 在多位函数中，求导数得到的是梯度，梯度是由编导数组成的向量，有着和x（参数）相同的形状，负梯度方向指向了函数下降最快的方向
3. 函数某点任意方向的斜率等于这一点上梯度和该点单位方向向量的点积

**计算梯度的方法：**

1. 有限差分法(the method of finite differences)，根据梯度的定义计算梯度，在W上加一个小的偏移量后计算loss值的变化，除以偏移量可近似得到梯度向量
2. 使用数值分析的方法计算梯度

**随机梯度下降**

不计算整个训练集的误差和梯度值，在每一次迭代中，选取一小部分训练样本组成minibatch（通常为2的n次幂）使用minibatch估算误差综合

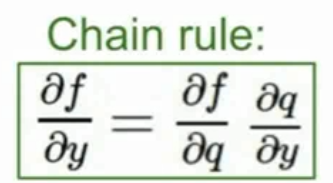
**图像的特征**

颜色直方图，方向梯度梯度图（将图像按八个像素区为八份，在八个像素区的每一部分计算像素值的主要边缘方向）、词袋bag of word（用特征向量表示一句话方法之一是计算不同词在这段话中出现的次数）

# 反向传播

计算任意复杂函数的解析梯度方法。计算图是指用来表示任意函数的图，图中的节点表示我们要执行的每一步的计算。

反向传播是链式法则的递归调用，从计算图的最后面向前计算所有的梯度



每个计算节点有一个本地梯度计算公式，将该公式与后向传播的梯度相乘可以得到该计算节点对应参数相对于整个f的梯度

**计算图节点详解**

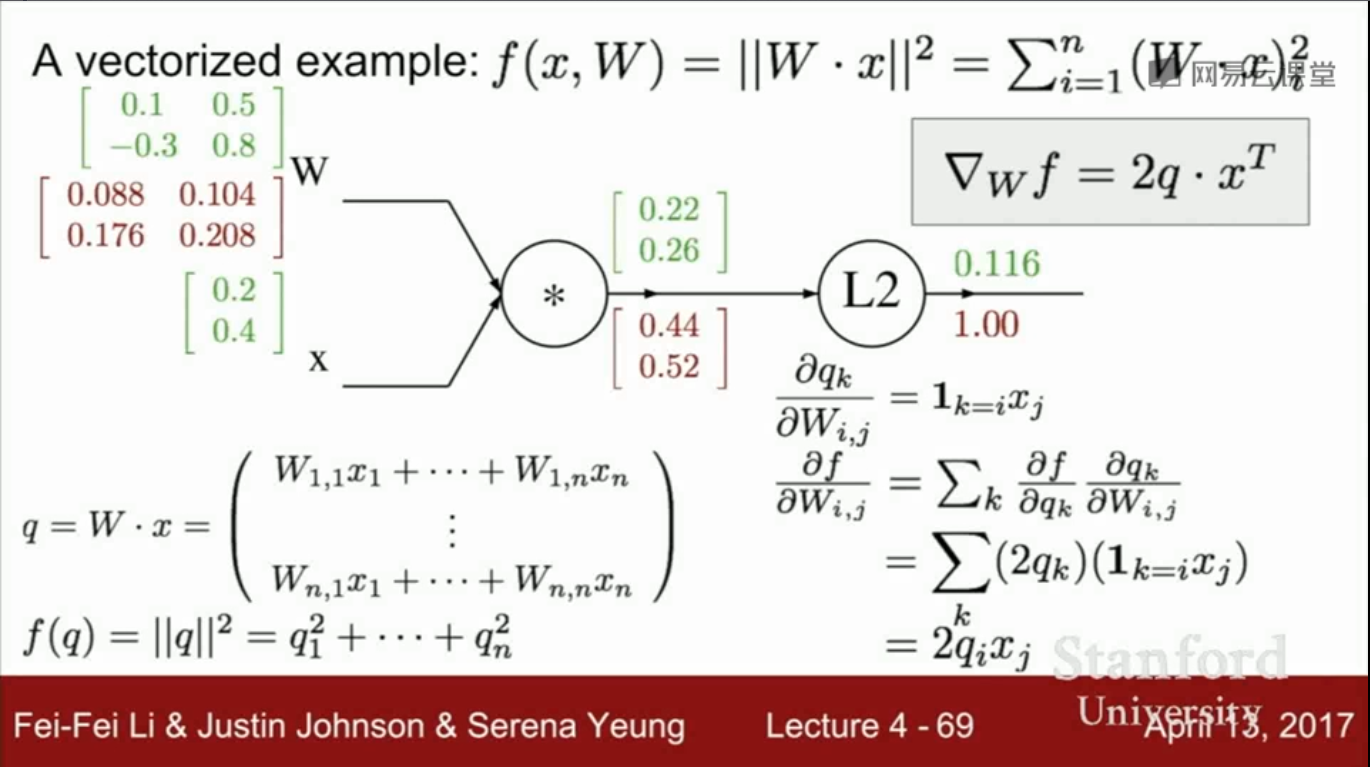
加法门是一个梯度分布器，将上游的梯度分发给下游的各个节点，max门是将上游的梯度只传递给下游其中一个节点，乘法门是一个梯度转换器，根据另一个分支的值对其进行缩放。

当一个节点的上游有多个运算节点时，梯度会在在这个节点累加

**变量是高维向量的情况**

这时梯度变为了雅克比矩阵，该雅克比矩阵中每一行都是输出对于输入求偏导的结果。但是由于第一个输入的参数只与第一输出有关，因此雅克比矩阵只是一个对角矩阵

例子



# 神经网络

神经网络是由简单函数构成的一组函数，用层次化的方法堆叠起来，来形成一个复杂的非线性函数。

权重矩阵的每一行相当于一个模板

# 神经网络