

神经网络入门

权重：一个较大的权重意味着神经网络认为这个输入比其它输入更重要，较小的权重意味着数据不是那么重要。

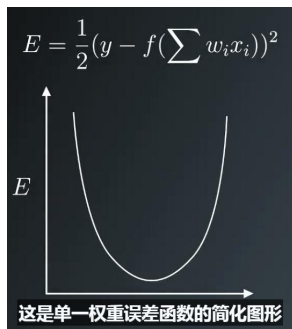
偏置：更多可能性返回某一个结果（或者说更倾向于某种结果）

激活函数：将求和的结果转换成某种输出信号

误差 (error)：一个普遍指标是误差平方和 sum of the squared errors (SSE)

$$E = \frac{1}{2} \sum_{\mu} (y^{\mu} - \hat{y}^{\mu})^2$$

使用平方而不使用绝对值的原因是：平方后误差大的得到的惩罚也越大，平方后误差小的得到的惩罚也越小；平方也可以简化后续的计算。1/2 也是为了简化后续步骤计算。



（我们的目的就是得到最小值时的权重值）

梯度下降：

- 误差只与权重有关，通过一步步改变权重，可以得到误差的最小值。
- 权重的更新与梯度的方向恰好相反。
- 学习速率用来控制梯度下降中更新步长的大小。

$$\frac{\partial E}{\partial w_i} = -(y - \hat{y}) f'(h) x_i$$
$$\Delta w_i = \eta (y - \hat{y}) f'(h) x_i$$

那么更新步长就等于学习速率 η 乘以预测差值

之前我们看到一个权重的更新可以这样计算：

$$\Delta w_i = \eta \delta x_i$$

这里 error term δ 是指

$$\delta = (y - \hat{y}) f'(h) = (y - \hat{y}) f'(\sum w_i x_i)$$

记住，上面公式中 $(y - \hat{y})$ 是输出误差， $f'(h)$ 是激活函数 $f(h)$ 的导函数，我们把这个导函数称做输出的梯度。