机器学习-第八讲课后作业

姓名：周宝航 学号：2120190442 专业：计算机科学与技术

一、问题描述

1. 考虑两层网络函数

$$y_k(x,w)=\sigma(\sum_{j=1}^{M}w_{kj}^{(2)}h(\sum_{i=1}^{D}w_{ji}^{(1)}x_i+w_{j0}^{(1)})+w_{k0}^{(2)})，$$

其中，隐含单元非线性激活函数是 Logistic sigmoid 函数

$$\sigma(a)=\{1+\exp(-a)\}^{-1}$$

证明：存在一个准确计算相同函数的等价网路，但该等价网络隐含单元激活函数是$\tanh(a)$函数

$$\tanh(a)=\frac{e^a-e^{-a}}{e^a+e^{-a}}。$$

2. 考虑多目标变量回归问题，假设目标标量对输入矢量 x 的条件分布满足高斯分布 $p(t\mid x,w)=N(t\mid y(x,w),\Sigma)$。其中 $y(x,w)$ 是网络在输入矢量 x 和权矢量 w 时的输出，Σ 是目标变量上高斯噪声的协方差矩阵。已知独立观测量集合$\{x,t\}$，如果假设Σ是不变和已知的，那么为了找到 w 的最大似然解必须最小化的误差函数是什么？现在假设Σ也是由数据决定的，写出Σ最大似然解的表达式。注意参数 w 和Σ的优化是耦合的。

二、基本思路

1. 假设我们已经有一个隐含层的激活函数为 logistic sigmoid 的神

经网络。我们考虑其隐含层的线性运算的输出结果为：w，那么经过激活函数后的输出值为：$\sigma(w)=\frac{1}{1+\exp(-w)}$。而为了保证替换激活函数为$\tanh(a)$后，神经网络的输出仍与之前一致，我们可以建立等式：$\sigma(w)=\tanh(w')$。其中w'表示激活函数为$\tanh(a)$时，神经网络的隐含层线性运算结果。如果证明两种激活函数对应的神经网络是等价的，那么只需要前面的方程中w'存在实数解即可。我们将等式展开得到：

$$\frac{1}{1+e^{-w}}=\frac{e^{w'}-e^{-w'}}{e^{w'}+e^{-w'}}$$

对上式化简可以得到：$w'=\frac{\ln(2e^{w}+1)}{2}$。由此，我们可以知道：对于相同输出值，$\tanh(a)$激活函数总有对应的输入值保证与 logistic sigmoid 函数相同。所以，证得：存在一个准确计算相同函数的等价网路，但该等价网络隐含单元激活函数是$\tanh(a)$函数。

2. 假设Σ是不变和已知的，那么寻找 w 的最大似然解必须最小化的误差函数是：$\frac{1}{2}\sum_{\mathrm{i=1}}^{\mathrm{N}}(t_i-y(x_i,w))^2$。我们考虑权值为$\hat{w}$的似然函数：

$$\mathrm{L}(\{\mathrm{x,t}\};\hat{\mathrm{w}},\Sigma)=\frac{1}{(2\pi)^{2/N}}\frac{1}{|\Sigma|^{1/2}}\exp\{-\frac{1}{2}(t-y(x,\hat{\mathrm{w}}))^{\mathrm{T}}\Sigma^{-1}(t-y(x,\hat{\mathrm{w}}))\}。$$

我们对上面的损失函数取对数并令协方差矩阵的偏导数为 0，解方程可以得到协方差矩阵的最大似然表达式为：

$$\Sigma=\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}}(t_i-y(x_i,\hat{w}))(t_i-y(x_i,\hat{w}))^T。$$