第13周作业展示

第一题

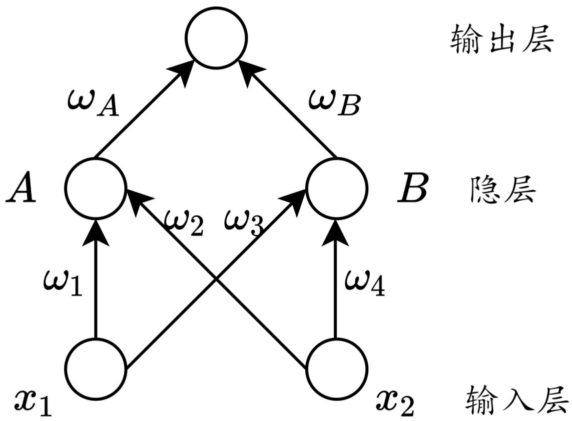
题目描述

试述将线性函数$f(x) = \omega^T x$用作神经元激活函数的缺陷

结果分析

神经网络是一个包含许多参数的数学模型，这个模型由若干个函数相互嵌套代入而得，而线性激活函数嵌套后依然是线性函数。使用线性函数作为激活函数时，无论是在隐藏层还是在输出层，其单元值都还是输入x的线性组合，若输出层也使用线性函数作为激活函数，那么无论多少层的神经网络都会退化成一个线性回归。因此，作为神经网络中必须要有非线性的激活函数。

以下图所示的二层感知机为例，举例说明，为了简化过程，我们忽略阈值θ：



对于神经元A和B，接受输入后经过激活函数处理后得到的输出为：

$$y_A = \omega_1 x_1 + \omega_2 x_2$$
$$y_B = \omega_3 x_1 + \omega_4 x_2$$

隐层经过激活函数处理后的输出作为输出层神经元的输入，输出层神经元经过激活函数处理后，最终输出为：

$$y = \omega_A y_A + \omega_B y_B$$

经过简单整理后，发现最终输出仍然为线性模型：

$$y = \omega_{10} x_1 + \omega_{20} x_2$$

如果采用累积误差逆传播算法，则累积误差为：

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

由于BP算法的目标是最小化训练集上的累计误差E，从上式可以看出，这与线性回归中最小化均方误差是一致的，此时就没有必要用神经网络进行训练了。

注意：经过同学指正，上述过程存在一些不严密的地方。严格地说，上述过程应该建立在激活函数为$f(\omega^T x) = \omega^T x$的前提下进行推导。但是由于只是举例线性函数的嵌套仍然是线性函数，因此对最终的结论论述没有影响。

第二题

题目描述

对于教材图5.7中的v_{ih}，试推导出BP算法中的更新公式。

结果分析

BP算法基于梯度下降，以目标的负梯度方向对参数进行调整，对于误差E_k，给定学习率η，有:

$$E_k = \frac{1}{2}\sum_{j=1}^{l}(\hat{y}_j^k - y_j^k)^2 \text{ , } \Delta v_{ih} = -\eta\frac{\partial E_k}{\partial v_{ih}}$$

由数学中的链式法则，有如下偏导关系:

$$\frac{\partial E_k}{\partial v_{ih}} = \frac{\partial E_k}{\partial b_h}\frac{\partial b_h}{\partial \alpha_h}\frac{\partial \alpha_h}{\partial v_{ih}}$$

根据上述式子，计算偏导前两项之积，有:

$$\begin{aligned} e_h &= -\frac{\partial E_k}{\partial b_h}\frac{\partial b_h}{\partial \alpha_h} \\ &= -\sum_{j=1}^{l}\frac{\partial E_k}{\partial \beta_j}\frac{\partial \beta_j}{\partial b_h}f'(\alpha_h - \gamma_h) \\ &= \sum_{j=1}^{l}\omega_{hj}g_j f'(\alpha_h - \gamma_h) \\ &= b_h(1-b_h)\sum_{j=1}^{l}\omega_{hj}g_j \text{ , 其中} g_j = -\frac{\partial E_k}{\partial \hat{y}_j^k}\frac{\partial \hat{y}_j^k}{\partial \beta_j} \end{aligned}$$

由于$\frac{\partial \alpha_h}{\partial v_{ih}} = x_i$,因此推导出更新公式:

$$\Delta v_{ih} = \eta e_h x_i$$

第三题

题目描述

给出教材(6.41)的完整的KKT条件。

结果分析

使用hinge损失，则最优化目标:

$$min \quad \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{m}max(0, 1 - y_i(\omega^T x_i + b))$$

引入松弛变量，给出的优化目标函数为:

$$min \quad \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{m}\xi_i$$

约束条件为:

$$\xi_i \geq 0 \quad i = 1, 2, \ldots, m$$
$$y_i(\omega^T x_i + b) \geq 1 - \xi_i$$

构造拉格朗日函数为:

$$L = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{m}\xi_i + \sum_{i=1}^{m}\alpha_i(1 - \xi_i - y_i(\omega^T x_i + b)) - \sum_{i=1}^{m}\mu_i\xi_i$$

对拉格朗日函数求偏导数得到:

$$\frac{\partial L}{\partial \omega}=\omega-\sum_{i=1}^{m}\alpha_i y_i x_i=0$$
$$\frac{\partial L}{\partial b}=\sum_{i=1}^{m}\alpha_i y_i=0$$
$$\frac{\partial L}{\partial \xi}=0\rightarrow C=\alpha_i+\mu_i$$

得到其的对偶问题：

$$max \quad \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$s.t.\sum_{i=1}^{m}\alpha_i y_i=0,0\le\alpha_i\le C,i=1,2,\ldots,m$$

在此求解过程中，应该满足KKT条件。

$$\text{拉氏条件}:\omega=\sum_{i=1}^{m}\alpha_i y_i x_i=0\text{；}\sum_{i=1}^{m}\alpha_i y_i=0\text{；}C=\alpha_i+\mu_i$$
$$\text{乘子条件}:\alpha_i\ge 0\quad;\quad\mu_i\ge 0$$
$$\text{原不等式约束}:1-\xi_i-y_i(w^T x_i+b)\le 0\text{；}\xi_i\ge 0$$
$$\text{互补松弛条件}:\alpha_i[y_i(\omega^T x_i+b)-1+\xi_i]=0\text{；}\mu_i\xi_i=0$$

第四题

题目描述

讨论线性判别分析与线性核支持向量机在何种条件下等价

结果分析

1.LDA与SVM的思想：LDA最大化类间散度，最小化类内散度；SVM找支持向量，并使其尽可能大。因此两算法的思想都不一样。

2.在控制支持向量不变的条件下，调整样例，SVM不会改变但LDA会改变，这进一步证明了两种算法是无法从条件层面比较的。

3.因此如果两种算法等价，那么一定是数据驱动的，而非是数学条件驱动的。SVM会解出ω，LDA会解出W_{c0}，两向量正交，此时SVM与LDA最多相差一个截距项

4.如果线性可分，那么只需满足上述条件，若线性不可分，则投射到高维后可以变得线性可分。

In [ ]: