第五章 神经网络作业

[redacted]

5.1: 试述将线性函数 $f(x)=w^Tx$ 用作神经元激活函数的缺陷。

答: 因无论在隐藏层还是在输出层使用激活函数之后都是关于 x 的线性组合，若仍使用线性函数 $f(x)=w^Tx$ 作为激活函数，整个过程都仍等价于线性回归，达不到"筛选"与"激活"的目的。

5.2: 试述使用图5.2(b)激活函数的神经元与对率回归的联系。

答: 相同点: 两者均使用了 sigmoid: $y=\frac{1}{1+e^{-x}}$ 替代函数

不同之处: 在对数几率回归中，当 $sigmoid(x)>0.5$ 时趋近为1 表示正例；反之为0，表示负例

在5.2(b)激活函数中，$sigmoid(x)$ 即为真实输出值。

5.3: 证明: $\Delta V_{ih}=-\eta\frac{\partial E_k}{\partial v_{ih}}=-\eta\frac{\partial E_k}{\partial \hat{y}_j^k}\cdot\frac{\partial \hat{y}_j^k}{\partial \beta_j}\cdot\frac{\partial \beta_j}{\partial b_h}\cdot\frac{\partial b_h}{\partial \alpha_h}\cdot\frac{\partial \alpha_h}{\partial v_{ih}}$

由于 $=\eta\sum_{j=1}^{l}g_j w_{hj}\cdot f'(\alpha_h-\gamma_h)\cdot x_i$ $\quad(b_h=f(\alpha_h-\gamma_h)$ γ_h 为阈值)

由于 $f'(\alpha_h-\gamma_h)=f(\alpha_h-\gamma_h)[1-f(\alpha_h-\gamma_h)]=b_h(1-b_h)$.

则 $\Delta V_{ih}=\eta\sum_{j=1}^{l}g_j w_{hj}\, b_h(1-b_h)\cdot x_i$

$=\eta\,\underline{b_h(1-b_h)\sum_{j=1}^{l}w_{hj}}\cdot x_i$

$=\eta\, e_h\, x_i$

5.4 解: 学习率太小，会使每次下降变化得越慢，训练耗费时间长，开销成本大。

学习率太大，会使每次下降变化幅度大，易在最优点附近来回振荡，可能达不到最优点。

第六章 支持向量机作业

6.1 证明：设超平面：$w^T\vec{x}+b=0$，点 $\vec{x}_1$ 为平面任一点，$\vec{x}_0$ 为点在超平面上的投影

则：$w^T\vec{x}_1+b=0$

则点 x_0 到平面的距离即为 $\vec{x}_1-\vec{x}_0$ 在 w^T 方向上的投影

$$r=\frac{|w^T(\vec{x}_1-\vec{x}_0)|}{||w^T||}=\frac{|w^T\vec{x}_1-w^T\vec{x}_0|}{||w||}=\frac{|-b-w^T\vec{x}_0|}{||w||}=\frac{|w^Tx_0+b|}{||w||}$$

证毕。

6.4 解：①首先应为二分类问题：因为虽然LDA可进行多分类，但SVM只能处理二分类问题。

②若为硬间隔，则需为线性可分；而软间隔则不用。

③当LDA方法求得的 $\vec{w}_1$ 与SVM求得的分隔超平面中的 w_2^T 相互垂直时

两者等价。（平面方向与法线方向同向）.

6.6 解：由于SVM目标是寻找一个分隔超平面获取"最大间隔"，在硬间隔划分中，若产生过多的噪声且噪声分布不均匀，则可能无法寻找到划分超平面；而对于软间隔，噪声可能落入最大间隔之中，成为支持向量，~~影响划分~~对分隔超平面的确定产生较大影响.

6.7 解：其KKT条件：

$$\begin{cases}
\mu_i\xi_i=0\\
\hat{\mu}_i\hat{\xi}_i=0\\
\xi_i\ge0,\ \hat{\xi}_i\ge0\\
\mu_i\ge0,\ \hat{\mu}_i\ge0\\
\alpha_i\ge0,\ \hat{\alpha}_i\ge0\\
\alpha_i\hat{\alpha}_i=0,\ \xi_i\hat{\xi}_i=0\\
\alpha_i(f(x_i)-y_i-\epsilon-\xi_i)=0\\
\hat{\alpha}_i(y_i-f(x_i)-\epsilon-\hat{\xi}_i)=0\\
f(x_i)-y_i-\epsilon-\xi_i\le0,\ y_i-f(x_i)-\epsilon-\hat{\xi}_i\le0
\end{cases}$$