分类器组合与集成习题

姓名： 甘云冲

学号： 2101213081

1.

(1)

写出拉格朗日函数如下：

$$L=\frac{1}{L}\sum_{j=1}^{L}\sum_{i=1}^{K}P(\omega_i|\boldsymbol{x})\ln\frac{P(\omega_i|\boldsymbol{x})}{P_j(\omega_i|\boldsymbol{x})}-\alpha\left(\sum_{i=1}^{k}P(\omega_i|\boldsymbol{x})-1\right)$$

求偏导：

$$\frac{\partial L}{\partial P(\omega_i|x)}=\frac{1}{L}\sum_{j=1}^{L}\ln P(\omega_i|\boldsymbol{x})+1-\ln P_j(\omega_i|\boldsymbol{x})-\alpha$$

$$\begin{aligned}\ln P(\omega_i|\boldsymbol{x})+1-\alpha&=\frac{1}{L}\sum_{j=1}^{L}\ln P_j(\omega_i|\boldsymbol{x})\\ \Rightarrow\ln P(\omega_i|\boldsymbol{x})&=\ln\left(\prod_{j=1}^{L}P_j(\omega_i|\boldsymbol{x})\right)^{1/L}+(\alpha-1)\\ P(\omega_i|\boldsymbol{x})&=C\left(\prod_{j=1}^{L}P_j(\omega_i|\boldsymbol{x})\right)^{1/L}\end{aligned}$$

其中C为归一化因子，由于$\sum_{i=1}^{K}P(\omega_i|\boldsymbol{x})=1$，应当有：

$$C=\frac{1}{\sum_{i=1}^{K}\left(\prod_{j=1}^{L}P_j(\omega_i|\boldsymbol{x})\right)^{1/L}}$$

得证原式。

(2)

$$
\begin{aligned}
&\min_{P(\omega_i|\boldsymbol{x})} \frac{1}{L}\sum_{j=1}^{L}\sum_{i=1}^{K} P_j(\omega_i|\boldsymbol{x}) \ln \frac{P_j(\omega_i|\boldsymbol{x})}{P(\omega_i|\boldsymbol{x})} \\
\Leftrightarrow &\max_{P(\omega_i|\boldsymbol{x})} \frac{1}{L}\sum_{j=1}^{L}\sum_{i=1}^{K} P_j(\omega_i|\boldsymbol{x}) \ln P(\omega_i|\boldsymbol{x}) \\
\Leftrightarrow &\max_{P(\omega_i|\boldsymbol{x})} \sum_{i=1}^{K}\left(\frac{1}{L}\sum_{j=1}^{L} P_j(\omega_i|\boldsymbol{x})\right) \ln P(\omega_i|\boldsymbol{x})
\end{aligned}
$$

可以知道：

$$
\sum_{i=1}^{K}\left(\frac{1}{L}\sum_{j=1}^{L} P_j(\omega_i|\boldsymbol{x})\right) \ln P(\omega_i|\boldsymbol{x}) \leq \sum_{i=1}^{K}\left(\frac{1}{L}\sum_{j=1}^{L} P_j(\omega_i|\boldsymbol{x})\right) \ln \left(\frac{1}{L}\sum_{j=1}^{L} P_j(\omega_i|\boldsymbol{x})\right)
$$

当：

$$
P(\omega_i|\boldsymbol{x}) = \left(\frac{1}{L}\sum_{j=1}^{L} P_j(\omega_i|\boldsymbol{x})\right)
$$

时候取到极大值。

2.

采用加权投票法，错误率为$0.2, 0.3, 0.4, 0.4, 0.4$的分类器的权重依次为$\frac{4}{9}, \frac{2}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}$。当错误的权重和$>$正确的权重和的时候，产生错误。

$$
P_{err} = 0.3 * 0.4^3 * 0.8 + 0.2 * (1 - 0.7 * 0.6^3) = 0.185 < 0.2
$$

3.

$$
\begin{aligned}
&\min_{\alpha,h} \sum_{i=1}^{n} [y_i - (f_{t-1}(\boldsymbol{x}_i) + \alpha h(\boldsymbol{x}_i))]^2 \\
\Leftrightarrow &\min_{\alpha,h} \sum_{i=1}^{n} [\beta_i^t - \alpha h(\boldsymbol{x}_i))]^2
\end{aligned}
$$

$\hat{\alpha}\hat{h}(\boldsymbol{x}_i)$为当前残差的最优拟合，所以更新过程为：

$$f_t(\boldsymbol{x}) = f_{t-1}(\boldsymbol{x}) + \hat{\alpha}_t \hat{h}_t(\boldsymbol{x})$$

这样可以在不影响前面分类器的情况下逐步增加分类器。

算法如下：

1. 初始化$f_0(\boldsymbol{x}) = 0$
2. 对$m = 1 \ldots M$：
    1. 计算$(\hat{\alpha}_t, \hat{h}_t) = \arg\min_{\alpha,h} \sum_{i=1}^{n} [\beta_i^t - \alpha h(\boldsymbol{x}_i))]^2$
    2. 更新$f_t(\boldsymbol{x}) = f_{t-1}(\boldsymbol{x}) + \hat{\alpha}_t \hat{h}_t(\boldsymbol{x})$