人工智能之机器学习

晚自习

主讲人：刘老师(GerryLiu)

课程要求

- 课上课下“九字”真言
  - 认真听，**善摘录，勤思考**
  - **多温故，乐实践**，再发散
- 四不原则
  - 不懒散惰性，不迟到早退
  - 不请假旷课，不拖延作业
- 一点注意事项
  - 违反“四不原则”，不推荐就业

课程内容

- Adaboost

AdaBoost算法原理

- 使下列公式达到最小值的α_m和G_m就是AdaBoost算法的最终解

$$loss\left(\alpha_m, G_m(x)\right) = \sum_{i=1}^{n} \overline{w}_{mi} e^{\left(-y_i \alpha_m G_m(x)\right)}$$

- G这个分类器在训练的过程中，是为了让误差率最小，所以可以认为G越好其实就是误差率越小。

$$G_m^*(x) = \min_{G_m(x)} \sum_{i=1}^{n} \overline{w}_{mi} I\left(y_i \neq G_m(x_i)\right)$$

$$\varepsilon_m = P\left(G_m(x) \neq y\right) = \sum_{i=1}^{n} \overline{w}_{mi} I\left(y_i \neq G_m(x_i)\right) = \sum_{y_i \neq G_m(x_i)} \overline{w}_{mi}$$

- 对于α_m而言，通过求导然后令导数为零，可以得到公式（log对象可以以e为底也可以以2为底）：

$$\alpha_m^* = \frac{1}{2} \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right)$$

扩展_AdaBoost算法子模型权重系数求解

$$\sum_{i=1}^{n}\overline{w}_{mi}=1 \qquad \varepsilon_m=\sum_{y_i\neq G_m(x_i)}\overline{w}_{mi}$$

$$
\begin{aligned}
loss(\alpha_m, G_m(x)) &= \sum_{i=1}^{n}\overline{w}_{mi}e^{(-y_i\alpha_m G_m(x))} \\
&= \sum_{y=G(x)}\overline{w}_{mi}e^{-\alpha_m} + \sum_{y\neq G(x)}\overline{w}_{mi}e^{\alpha_m} \\
&= \sum_{y=G(x)}\overline{w}_{mi}e^{-\alpha_m} + \varepsilon_m e^{\alpha_m} \\
&= \sum_{y=G(x)}\overline{w}_{mi}e^{-\alpha_m} + \varepsilon_m e^{\alpha_m} + \sum_{y\neq G(x)}\overline{w}_{mi}e^{-\alpha_m} - \sum_{y\neq G(x)}\overline{w}_{mi}e^{-\alpha_m} \\
&= \sum_{i=1}^{n}\overline{w}_{mi}e^{-\alpha_m} + \varepsilon_m e^{\alpha_m} - \varepsilon_m e^{-\alpha_m} \\
&= e^{-\alpha_m} + \varepsilon_m e^{\alpha_m} - \varepsilon_m e^{-\alpha_m}
\end{aligned}
$$

扩展_AdaBoost算法子模型权重系数求解

$$loss = e^{-\alpha_m} + \varepsilon_m e^{\alpha_m} - \varepsilon_m e^{-\alpha_m} \qquad \frac{\partial loss}{\partial \alpha_m} = -e^{-\alpha_m} + \varepsilon_m e^{\alpha_m} + \varepsilon_m e^{-\alpha_m}$$

$$\Rightarrow \frac{\partial loss}{\partial \alpha_m} = 0 \Rightarrow -e^{-\alpha_m} + \varepsilon_m e^{\alpha_m} + \varepsilon_m e^{-\alpha_m} = 0$$

$$\Rightarrow (\varepsilon_m - 1)e^{-\alpha_m} + \varepsilon_m e^{\alpha_m} = 0 \Rightarrow \varepsilon_m e^{\alpha_m} = (1-\varepsilon_m)e^{-\alpha_m}$$

$$\Rightarrow \frac{e^{\alpha_m}}{e^{-\alpha_m}} = \frac{(1-\varepsilon_m)}{\varepsilon_m} \Rightarrow e^{2\alpha_m} = \frac{(1-\varepsilon_m)}{\varepsilon_m}$$

$$\Rightarrow \ln e^{2\alpha_m} = \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right) \Rightarrow 2\alpha_m = \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right)$$

$$\Rightarrow \alpha_m = \frac{1}{2}\ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right)$$

GBDT讲解

- 最优的算法模型实际上就是预测值F(x)等于实际值y。也就是在我们模型训练的过程中，我们是希望模型的预测值和实际值是完全相同的，而且模型训练的方向也是朝着这个方向训练。