2.1.1

Parent entropy: $H(D) = -\left(\frac{3}{10}\log\frac{3}{10} + \frac{2}{10}\log\frac{2}{10} + \frac{3}{10}\log\frac{3}{10} + \frac{2}{10}\log\frac{2}{10}\right) = 1.97$

split by 天气: 晴朗 $H_1(D_1) = -\left(\frac{2}{5}\log\frac{2}{5} + \frac{1}{5}\log\frac{1}{5} + \frac{2}{5}\log\frac{2}{5}\right) = 1.52$

多云 $H_1(D_2) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 1$

下雨 $H_1(D_3) = -\left(\frac{3}{3}\log\frac{3}{3}\right) = 0$

$IG_1 = H(D) - \frac{5}{10}H_1(D_1) - \frac{2}{10}H_1(D_2) - \frac{3}{10}H_1(D_3) = 1.97 - \frac{1}{2}\times 1.52 - \frac{1}{5}\times 1 - \frac{3}{10}\times 0$
$= 1.01$

split by 温度: 高 $H_2(D_1) = -\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) = 0.92$

中 $H_2(D_2) = -\left(\frac{2}{4}\log\frac{2}{4} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4}\right) = 1.5$

低 $H_2(D_3) = -\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) = 0.92$

$IG_2 = H(D) - \frac{3}{10}H_2(D_1) - \frac{4}{10}H_2(D_2) - \frac{3}{10}H_2(D_3) = 1.97 - \frac{3}{10}\times 0.92\times 2 - \frac{2}{5}\times 1.5$
$= 0.82$

split by 湿度: 低: $H_3(D_1) = -\left(\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right) = 0.97$

高: $H_3(D_2) = -\left(\frac{2}{5}\log\frac{2}{5} + \frac{3}{5}\log\frac{3}{5}\right) = 0.97$

$IG_3 = H(D) - \frac{5}{10}H_3(D_1) - \frac{5}{10}H_3(D_2) = 1.97 - \frac{1}{2}\times 2\times 0.97 = 1$

split by 风力: 强: $H_4(D_1) = -\left(\frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3}\right) = 1.58$

弱: $H_4(D_2) = -\left(\frac{3}{7}\log\frac{3}{7} + \frac{1}{7}\log\frac{1}{7} + \frac{1}{7}\log\frac{1}{7} + \frac{2}{7}\log\frac{2}{7}\right) = 1.84$

$IG_4 = H(D) - \frac{3}{10}H_4(D_1) - \frac{7}{10}H_4(D_2) = 1.97 - \frac{3}{10}\times 1.58 - \frac{7}{10}\times 1.84$
$= 0.21$

用天气作为根结点划分信息增益最大

2.1.2

① 天气 下雨 $\longrightarrow$ 看电影

② 天气 晴朗 $\longrightarrow$ 去徒步 2
去野餐 2
去博物馆 1

split by 温度: 高 $H_2'(D_1) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 1$

中 $H_2'(D_2) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 1$

低 $H_2'(D_3) = -\log 1 = 0$

$IG_2' = H_1(D_1) - \frac{2}{5}H_2'(D_1) - \frac{2}{5}H_2'(D_2) - \frac{1}{5}H_2'(D_3) = 1.52 - \frac{2}{5}\times 1 - \frac{2}{5} = 0.72$

split by 湿度: 低 $H_3'(D_1) = -\left(\frac{2}{4}\log\frac{2}{4} + \frac{2}{4}\log\frac{2}{4}\right) = 1$

高 $H_3'(D_2) = -\log 1 = 0$

$IG_3' = H_1(D_1) - \frac{4}{5}H_3'(D_1) - \frac{1}{5}H_3'(D_2) = 1.52 - \frac{4}{5}\times 1 = 0.72$

split by 风力: 强 $H_4'(D_1) = -\log 1 = 0$

弱 $H_4'(D_2) = -\left(\frac{2}{4}\log\frac{2}{4} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4}\right) = 1.5$

$IG_4' = H_1(D_1) - \frac{1}{5}H_4'(D_1) - \frac{4}{5}H_4'(D_2) = 1.52 - \frac{4}{5}\times 1.5 = 0.32$

按湿度与温度分，信息增益同样大，我选湿度

晴朗 — 低湿 去徒步 2
去野餐 2
晴朗 — 高湿 去博物馆

晴朗低湿时:

split by 温度: 高 $H_2''(D_1) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 1$

中 $H_2''(D_2) = -\log 1 = 0$

低 $H_2''(D_3) = -\log 1 = 0$

$IG_2'' = 1 - \frac{2}{4}\times 1 = 0.5$

split by 风力：强 $H_4''(D_1) = -\log 1 = 0$

弱 $H_4''(D_2) = -(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}) = 0.92$

$IG_4'' = 1 - \frac{3}{4}\times 0.92 = 0.31$

温度信息增益更大

晴朗 — 低湿 — 高温 1去徒步 1去野餐 最后按风力分
— 中温 去野餐
— 低温 去徒步

③天气多云 → 1去徒步
1去博物馆

split by 湿度：低 $H_3'''(D_1) = -\log 1$

高 $H_3'''(D_2) = -\log 1$

$IG_3''' = 1-0=1$

split by 温度：高 $H_2'''(D_1) = -\log 1$

中 $H_2'''(D_2) = -\log 1$

低 $H_2'''(D_3) = 0$

$IG_2''' = 1-0=1$

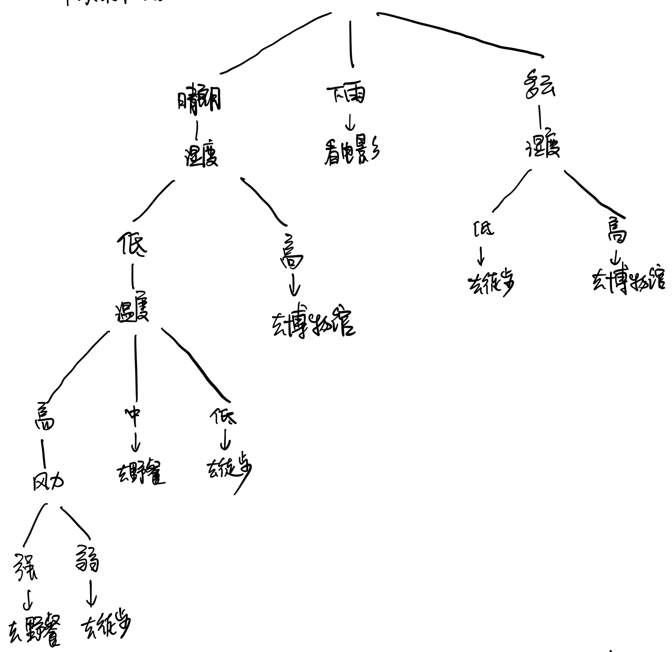
split by 风力：强：$H_4'''(D_1) = -\log 1$

弱：$H_4'''(D_2) = -\log 1$

$IG_4'' = 1-0=1$

三个features都可以用，我选 湿度

一个决策树为：



2.1.3 跟据以上决策树，晴朗，高温，高湿，弱风 去博物馆

2.2.4 二分类问题我认为泛化能力会更好一些，而回归树提供了清晰的决策边界。在这里，我发现树的深度对该问题影响不大

2.3.1 1颗树中任意样本被至少选1次的概率 $1-(1-\frac{1}{100})^{100}$

所有树中都未被选中的概率 $((1-\frac{1}{100})^{100})^{10} = (1-\frac{1}{100})^{1000}$

在至少1颗树中被至少选中1次的概率 $1-(1-\frac{1}{100})^{1000}$

2.3.2 任意一特征单次不被选中的概率 $\frac{C_{19}^{10}}{C_{20}^{10}}$

任意一特征被选中的概率 $1-\frac{C_{19}^{10}}{C_{20}^{10}}$

整个过程任一特征至少被选中1次的概率 $1-\left(\frac{C_{19}^{10}}{C_{20}^{10}}\right)^{10}$

3.1

1. 令 $f_0(\mathbf{x}) = 0$。
2. For t=1 to T：
   (a) 计算在各个数据点上的梯度 $\mathbf{g}_t = \left(\frac{\partial}{\partial \widehat{\mathbf{y}}_i}\ell(\mathbf{y}_i, \widehat{\mathbf{y}}_i)|_{\widehat{\mathbf{y}}_i = f_{t-1}(\mathbf{x}_i)}\right)_{i=1}^n$。
   (b) 根据 $-\mathbf{g}_t$ 拟合一个回归模型，$h_t = \arg\min_{h\in\mathcal{F}}$ $\sum_{i=1}^{n}\ell(y_i, f_{t-1}(x_i)+\alpha_t h_t(x_i))$。
   (c) 选择合适的步长 α_t, 最简单的选择是固定步长 $\eta \in (0, 1]$。
   (d) 更新模型，$f_t(\mathbf{x}) =$ $f_{t-1}(x) - \eta g_t$。

请完成以下题目：

1. 完成上述算法中的填空 (5pt)。

3.2

$$g_t = \left(\frac{\partial}{\partial \hat{y}_i}\frac{1}{2}(y_i-\hat{y}_i)^2 \Big|_{\hat{y}_i = f_{t-1}(x_i)}\right)_{i=1}^n$$

$$= -(y_i - f_{t-1}(x_i))_{i=1}^n$$

$$h_t = (y_i - f_{t-1}(x_i))_{i=1}^n$$

3.3

$$g_t = \left(\frac{\partial}{\partial \hat{y}_i}\ln(1+e^{-y_i\hat{y}_i}) \Big|_{\hat{y}_i = f_{t-1}(x_i)}\right)_{i=1}^n$$

$$= -\left(\frac{y_i e^{-y_i f_{t-1}(x_i)}}{1+e^{-y_i f_{t-1}(x_i)}}\right)_{i=1}^n = \left(-\frac{y_i}{1+e^{y_i f_{t-1}(x_i)}}\right)_{i=1}^n$$

$$h_t = \left(\frac{y_i}{1+e^{y_i f_{t-1}(x_i)}}\right)_{i=1}^n$$

3.6 在 L2 loss 和 logistic loss 的2分类问题中，弱学习器数量对分类结果影响不大，

总体上 logistic loss 处理2分类问题更准确

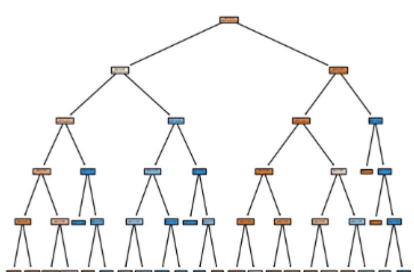
在回归问题上，发现弱学习器发挥很大作用，数量越多越准确.

4.3 发现选取树的数量为100，深度为10，特征数量为10比较好

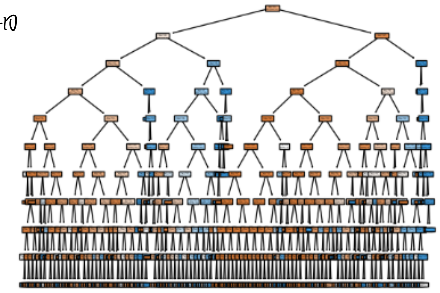
4.2 & 4.4

深度越深，树越长，深度为10时，准确率较大 (0.84705)

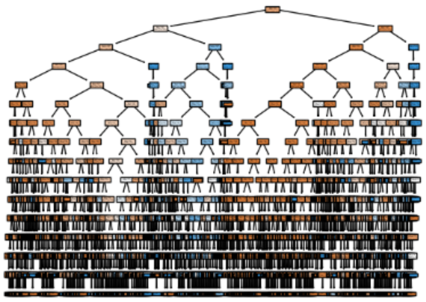
深度=5



深度=10



深度=15



4.5 通过调整学习步长 {0.01, 0.1, 0.5} 与树深度 {5, 10}

发现步长 0.1, 深度10时 准确率较高