hw2

第一题

（1）错误。L1 正则化的限定区域为矩形，更可能得到更为稀疏的参数向量。而 L2 正则化的限定区域为圆形，不易产生零解。

（2）错误。交叉验证只能减缓过拟合，而不能防止。如果数据集中特征和分类的分布不均匀，或者所选取的模型偏差较重，都会导致模型的过拟合。

（3）正确。样本量越大，越能避免少量样本分布不均匀导致分类错误的概率。

（4）错误。与用于分类的决策树相比，用于回归的决策树的差别在于：1. 设计算法找到连续变量的最佳切分点；2. 输出空间为单元内均值。

（5）正确。Bootstrap 只能降低方差。从数据中随机取样并不能改变数据的分布。

第二题

(1)

$$V^{\pi}(s)=\mathbb{E}_{\pi}[G_t|S_t=s]$$

(2)

$$V^{\pi}(s)=\mathbb{E}_{\pi}[R_{t+1}+\gamma V^{\pi}(S_{t+1})|S_t=s]$$

(3)

$$\begin{aligned}V_1^{\pi}(A)&=\mathbb{E}_{\pi_0}[R_1+\gamma V_0^{\pi_0}(S_1)|S_0=A]\\&=R_{ab}+\gamma V_0^{\pi_0}(B)\\&=-4+0.5\times 0\\&=-4\end{aligned}$$

$$\begin{aligned}V_1^{\pi}(B)&=\mathbb{E}_{\pi_0}[R_1+\gamma V_0^{\pi_0}(S_1)|S_0=B]\\&=0.5\times(R_{ba}+\gamma V_0^{\pi_0}(A))+0.5\times(R_{bc}+\gamma V_0^{\pi_0}(C))\\&=0.5\times(1+0.5\times 0)+0.5\times(2+0.5\times 0)\\&=1.5\end{aligned}$$

$$\begin{aligned}V_1^{\pi}(C)&=\mathbb{E}_{\pi_0}[R_1+\gamma V_0^{\pi_0}(S_1)|S_0=C]\\&=0.5\times(R_{cb}+\gamma V_0^{\pi_0}(B))+0.5\times(R_{ca}+\gamma(0.25V_0^{\pi_0}(C)+0.75V_0^{\pi_0}(A)))\\&=0.5\times(0+0.5\times 0)+0.5\times(8+0.5\times(0.25\times 0+0.75\times 0))\\&=4\end{aligned}$$

第三题

利用 hinge 损失，即：

$$l(f(\boldsymbol{x},y))=\max\{0,1-yf(\boldsymbol{x})\}$$

对于给定的点 (x_i,y_i)，优化目标等价为：

$$f(\boldsymbol{w};i)=\frac{1}{2}\|\boldsymbol{w}\|^2+C\gamma_i(\max\{0,1-y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)\})$$

其次梯度为：

$$\bigtriangledown_i = \boldsymbol{w} - I[y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) < 1]C\gamma_i y_i \boldsymbol{x_i}$$

其中，当 $y_i\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle < 1$ 时 $I = 1$ ，否则 $I = 0$ 。

因此算法采用迭代式算法，在第 t 轮迭代时选取 $\boldsymbol{x}_{it}$ ，并对 $\boldsymbol{w}$ 进行更新：

$$\boldsymbol{w}_{t+1} \leftarrow (1 - \eta_t)\boldsymbol{w}_t + \eta_t I[y_{it}(\boldsymbol{w} \cdot \boldsymbol{x}_{it} + b) < 1]C\gamma_{it} y_{it} \boldsymbol{x_i t}$$

其中 η_t 为第 t 次迭代的步长。

伪代码：

$$
\begin{aligned}
&\text{Input: } X,\ Y,\ T,\ \eta \\
&\text{Initialize: Set } \boldsymbol{w}_1 = 0 \\
&\text{For t=1 to T} \\
&\quad \text{Choose } \boldsymbol{x}_{it} \text{ randomly from } X \text{ randomly} \\
&\quad \text{Select } y_{it} \text{ from } Y \\
&\quad \text{If } y_i\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle < 1 \\
&\qquad \boldsymbol{w}_{t+1} \leftarrow (1 - \eta_t)\boldsymbol{w}_t + \eta_t C\gamma_{it} y_{it} \boldsymbol{x_{it}} \\
&\quad \text{Else} \\
&\qquad \boldsymbol{w}_{t+1} \leftarrow (1 - \eta_t)\boldsymbol{w}_t \\
&\text{Output: } \boldsymbol{w}_{t+1}
\end{aligned}
$$