學號：410551027
姓名：江衍涵

Part 2：

1. From pruning criterion is

$$C(T) = \sum_{\tau=1}^{|T|} Q_T(T) + \lambda|T|,$$

$$Q_\tau(T) = \sum_{k=1}^{K} P_{\tau k} \ln P_{\tau k}$$

We evalute the cross-entropy

$$C_{x\,ent}(T_A) = -2\left(\frac{100}{400}\ln\frac{100}{400} + \frac{300}{400}\ln\frac{300}{400}\right) + 2\lambda \simeq 1.12 + 2\lambda$$

$$C_{x\,ent}(T_B) = -\frac{400}{400}\ln\frac{400}{400} - \frac{200}{400}\ln\frac{200}{400} - \frac{0}{400}\ln\frac{0}{400} - \frac{200}{400}\ln\frac{200}{400} + 2\lambda$$

$$\simeq 0.69 + 2\lambda$$

Finally, From pruning criterion for the Gini index is

$$C(T) = \sum_{\tau=1}^{|T|} Q_\tau(T) + \lambda|T|,$$

$$Q_\tau(T) = \sum_{k=1}^{K} p_{\tau k}\,(1 - p_{\tau k})$$

$$C_{Gini}(T_A) = 2\left[\frac{300}{400}\left(1 - \frac{300}{400}\right) + \frac{100}{400}\left(1 - \frac{100}{400}\right)\right] + 2\lambda = \frac{3}{4} + 2\lambda$$

$$C_{Gini}(T_B) = \frac{400}{400}\left(1 - \frac{400}{400}\right) + \frac{200}{400}\left(1 - \frac{200}{400}\right) + \frac{0}{400}\left(1 - \frac{0}{400}\right)$$

$$+ \frac{200}{400}\left(1 - \frac{200}{400}\right) + 2\lambda = \frac{1}{2} + 2\lambda$$

Thus we see that, while both trees have the same missclassification rate, T_B is better in terms of cross-entropy as well as Gini index.

2.

Derivative =

$$E_{x,t}\left[e^{-ty(x)}\right] = \sum_t \int e^{-ty(x)} p(t|x)\,p(x)\,dx$$

$$\frac{\delta}{\delta y(x)} E_{x,t}\left(e^{-ty(x)}\right) = -\sum t\, e^{(-ty(x))}\, p(t|x)\,p(x)$$

$$\Rightarrow \left\{e^{y(x)}\, p(t=-1|x) - e^{-y(x)}\, p(t=+1|x)\right\} p(x) \quad \text{— (1)}$$

Make zero into the equation (1),

$$\Rightarrow y(x) = \frac{1}{2} \ln \frac{p(t=1|x)}{p(t=-1|x)}$$