1. **Generalization of Breiman’s theorem to splits in $k$ parts.**

Breiman’s theorem applies only to the problem of partitioning the set $I$ into two parts. Let’s consider the generalization of Breiman’s theorem for the case when we want to partition in $k$ parts instead.

**Theorem 1.** Let $I$ be a finite set, $q_i, r_i, i \in I$ be positive quantities and $\Phi(x)$ a concave function. For $I_1, \ldots, I_k$ a partitioning of $I$, an optimum of the problem

$$
\arg\min_{I_1, \ldots, I_k} \sum_{j=1}^{k} \left[ \sum_{i \in I_j} q_i \Phi \left( \frac{\sum_{i \in I_j} q_i r_i}{\sum_{i \in I_j} q_i} \right) \right]
$$

has the property that:

$$
\forall j < j', i \in I_j, \forall i' \in I_{j'}, r_i < r_{i'}
$$

(a) Explain how this theorem can be used to find splits in $k$ parts for building Breiman-like classification trees.

(b) Give an efficient algorithm (as efficient as possible) that uses the theorem to find the split.
(c) Prove the theorem.

Hint: The trick is to put most of the burden on Breiman’s theorem. The theorem effectively forces the sets $I_j$ to appear in order and have all elements ordered by quantity $r_i$. A method to do the proof is to consider an arbitrarily ordered set with an arbitrary partitioning and to show that, by sorting the elements, the value of the criterion gets better and better. You have to find a way to deal with items that have the same value for $r_i$ since they can create problems.

2. Generalization of Breiman’s theorem to splits in $k$ parts and $l$ class labels.

Breiman’s theorem applies only if there are only two class labels and a split in two. In the previous problem we looked at the generalization for splits in $k$ parts. It turns out that a generalization to $l$ class labels and $k$ parts is possible and it is provided by the following theorem:

**Theorem 2 (Burshtein et al. 1992).** Let $X$ be a random variable with values in $\mathcal{X}$, $\mathcal{U} \subset \mathbb{R}^n$ convex set, $Y : \mathcal{X} \to \mathcal{U}$ a measurable function so $Y$ is a random variable on $\mathcal{X}$ and $\mathcal{C} = \{1, \ldots, k\}$. Let $\phi : \mathcal{C} \times \mathcal{U} \to \mathbb{R}$ be a concave function in it’s second argument, $C : \mathcal{X} \to \mathcal{C}$ a measurable partitioning function and define $\Psi(C)$ as:

$$\Psi(C) = \sum_{c \in \mathcal{C}} P[C^{-1}(c)] \phi \left( c, \frac{\int_{C^{-1}(c)} Y(x) P(dx)}{P[C^{-1}(c)]} \right)$$

Then, for any partitioning function $C$ there exists a partitioning function over $\mathcal{U}$, $\tilde{C} : \mathcal{U} \to \mathcal{C}$ such that $\Psi(\tilde{C}(Y)) \leq \Psi(C)$ and $\tilde{C}^{-1}(c)$ is convex for all $c \in \mathcal{C}$.

(a) Show that the theorem indeed generalizes Breiman’s theorem (by showing how to instantiate it so that it is equivalent to Breiman’s theorem).

(b) Show how this theorem can be used to solve the problem of efficiently finding the optimal partitioning in $k$ parts when there are $l$ class labels.

(c) Estimate the complexity of the algorithm that would use the theorem to find such a partitioning.
You might want to use the following facts:

- A convex set is a set that has the property that, if it contains any two points $A$ and $B$ than it contains all the points on the line between $A$ and $B$.

- Convexity implies compactness, i.e. a convex set does not have holes. An example of convex set are intervals on the real line.

- $C^{-1}(c) = \{ x \in \mathcal{X} | C(x) = c \}$ If you associate $c$ with a partitioning than $C^{-1}(c)$ contains the elements in the partitioning (just a fancy way to write this).

- If $\mathcal{X}$ is discrete, than $\int_{C^{-1}(c)} Y(x) P[dx]$ is simply $\sum_{x, C(x) = x} Y(x) P[\{x\}]$

- Since $U$ is a subset of $\mathbb{R}^n$, $Y(x)$ has to be a random vector.

- Think about what the convexity of $\tilde{C}^{-1}(c)$ means

- In general be very careful since the notation is very heavy (a lot of symbols look alike but they are not the same thing).