Weakest link lagrangian derivation

To calculate the **error rate** in a tree T :

$$R(T) = \sum_{m=1}^{|\tilde{T}|} N_m Q_m$$

- N_m number of **observation** in a region m
- $Q_m(T)$ is the **impurity term** in **regression** RSS and in **classification** $misclassification$

Our goal is the **minimize** $R(T)$ alone results in a large tree with many **regions** but that's will just result in an overfit tree(maximal tree).

That's why we put a condition to control the **model complexity** which is related to number of **terminal nodes** formally written:

$$\min_T R(T) \quad \text{s.t. } |\tilde{T}| \leq k$$

Using the **Lagrangian** relaxation for the constraint :

$$\mathcal{L}(T, \alpha) = R(T) + \alpha(|\tilde{T}| - K)$$

- The α is a **penalty parameter** for tree size, the larger the simpler the tree
- α is the **parameter** for the **trade-off** between **impurity** and **leaves added**
- The **Lagrangian** form simply removes the constraint and add it to the **objective function** with the **lagrangian Multiplier** α
- Check Convex Optimization, Lagrange Multipliers for more details about the topic.

Since K is a constant it can be written as :

$$\mathcal{L}(T, \alpha) = R(T) + \alpha|\tilde{T}|$$

Which looks just as the **Cost Complexity** :

$$C_\alpha(T) = \sum_{m=1}^{|\tilde{T}|} N_m Q_m(T) + \alpha|\tilde{T}|$$

With $R(T) = \sum_m N_m Q_m$.

Conclusion : The cost complexity is lagrangian relaxation for a constrained problem, and pruning solves the constrained optimization without combinatorial search