# Decision Trees

## Useful formulas

- Entropy of a set S with elements from C classes: $H(S) = \sum_{i=1}^{C} - p_i \log p_i$.
- Gini impurity of a set S with elements from C classes: $G(S) = 1 - \sum_{i=1}^{C} p_i^2$
- Gini/Information gain: $G(S, F) = M(S) - \sum_{f \in values(F)} \frac{|S_f|}{|S|} M(S_f)$, where M is either the Gini impurity or the Entropy.

## Questions

1. What is the entropy of a dataset? How do you compute it?
2. How does the algorithm ID3 decide what the next feature to split on is?
3. What does ID3 do when there are no more features left to split on?
4. What type of data can decision trees classify which MLPs cannot?
5. What is a random forest? What are the sources of randomness that diversify the trees in the forest?
6. What is the main difference between the CART and the ID3 algorithms?
7. Consider the following dataset, where data have two features, each of which has three values (A, B, or C), and the last element is the class: <A, B,0>, <A, C,1>, <A, B,0>, <B, B,0>, <B, B, 0>, <B, C,1>, <C, A,1>, <C, B,1>, <C, B,1>, <C, C,0>. Construct a decision tree on the dataset with ID3.
8. We want to learn a classifier for car diagnosis. The classes are: OK (O); go to a garage (G); severe failure, don't drive (F). The features are: makes a strange noise (N) or not (nN); emits black smoke (S) or not (nS); going straight, the car drifts on a side (D), or doesn't (nD). We ask a mechanic, and build the following (very extensive) dataset: <N, nS,nD,G>, <nN,nS,nD,O>, <nN,S,nD,F>, <nN,S,D,F>, <N,S,nD,F>, <nN,nS,D,G>, <N,nS,D,G>, <N,nS,nD,G>. Construct a decision tree on the dataset with ID3. My car makes a strange noise, what should I do?
9. Same as the last two questions, but with CART.