Stanford University

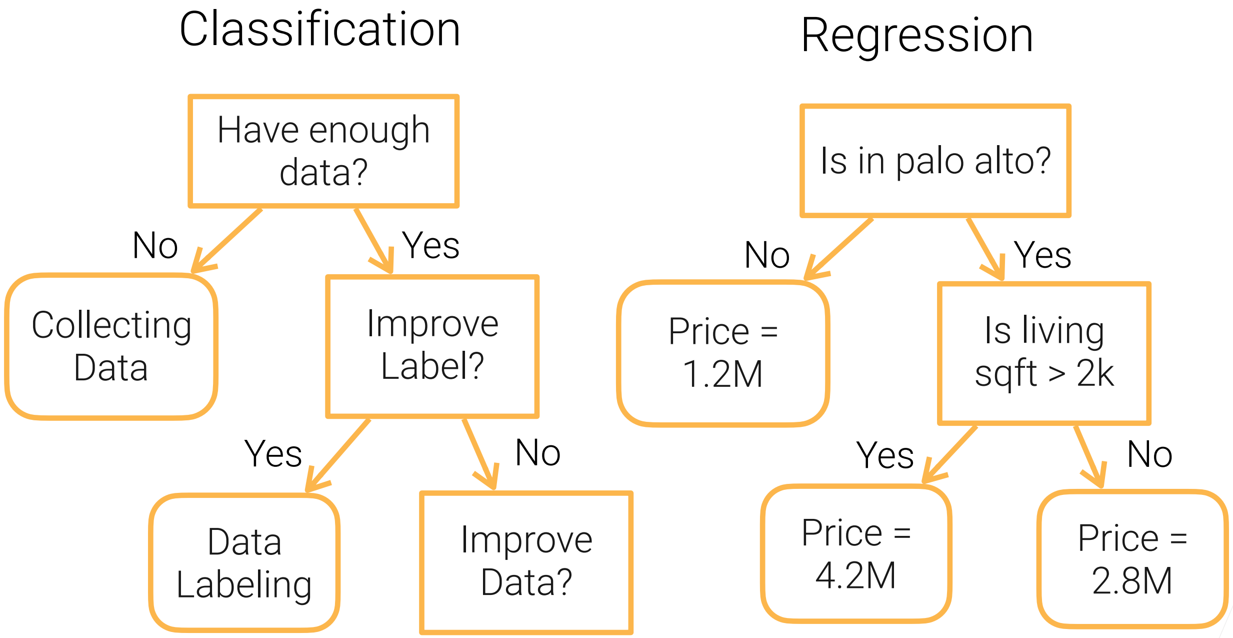
CS 329P : Practical Machine Learning (2021 Fall)

3.2 Decision Trees

Qingqing Huang, Mu Li, Alex Smola

https://c.d2l.ai/stanford-cs329p

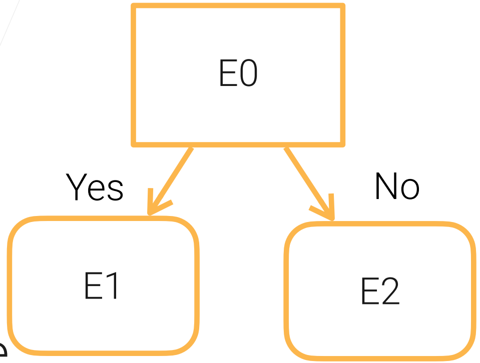
Decision Trees



- Explainable
- Handle both numerical and categorical features without preprocessing

Building Decision Trees

- Use a top-down approach,
  staring from the root node with the set of all features
- At each parent node, pick a feature to split the examples
  - Feature selection criteria
    - Maximize variance reduction for continuous target
    - Maximize Information gain (1 - entropy) for categorical target
    - Maximize Gini impurity $= 1 - \sum_{i=1}^{n} p_i^2$ for categorical target
  - All examples are used for feature selection at each node



Limitations of decision Trees

- Over-complex trees can overfit the data
  - Limit the number of levels of splitting,
  - Prune branches
- Sensitive to data
  - Changing a few examples can cause picking different features that lead to a different tree
  - Random forest
- Not easy to be parallelized in computing

Random Forest

- Train multiple decision trees to improve robustness
  - Trees are trained independently in **parallel**
  - Majority voting for classification, average for regression
- Where is the randomness from?
  - Bagging: randomly sample training examples with replacement
    - E.g. [1,2,3,4,5] → [1,2,2,3,4]
  - Randomly select a subset of features

Summary

- Decision tree: an explainable model for classification/regression
- Easy to train and tune, widely used in industry
- Sensitive to data
  - Ensemble can help (more on bagging and boosting latter)