LDA, Decision Trees, and Extra Trees

on the MNIST and Yale B Datasets

Kudiyar Orazymbetov, Nico Casale

NCSU

2017/04/23

outline

motivation

-
- applications:

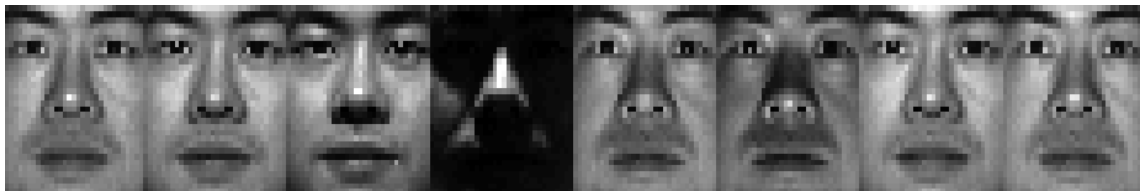
datasets

Modified Nat'l Institute of Standards and Technology (MNIST) database

- source: Yann LeCun et al. [1]
-

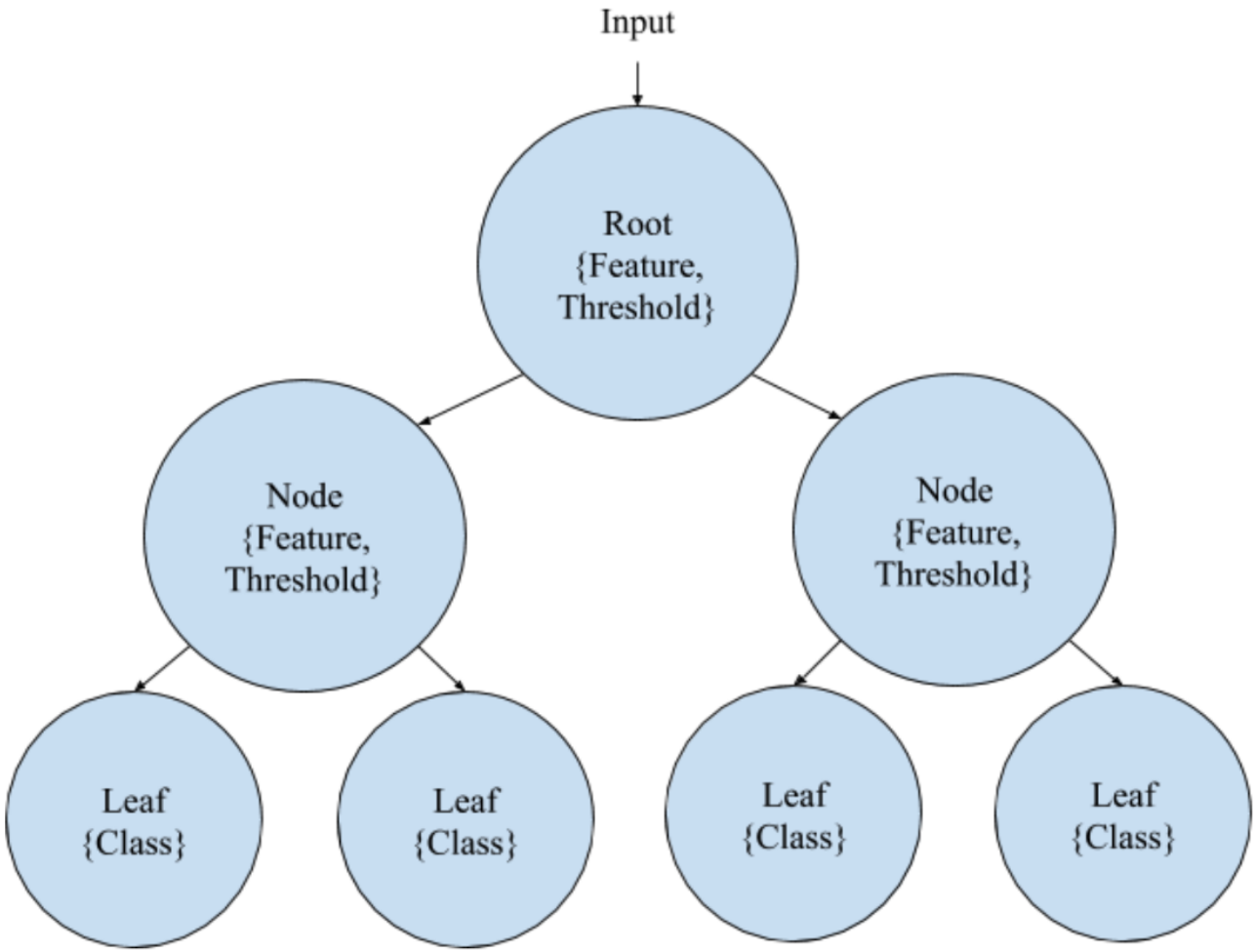
Yale Exended Face Database B

- source: Yale University [2]
-



progress

decision trees



training decision trees

training algorithm:

1. check stopping conditions
   - 1.1 no more features
   - 1.2 set is smaller than `minLeaf`
   - 1.3 all samples in the same class
   - 1.4 no feature improves information gain (IG)
2. iterate over each available feature, perform a line search to approximate the highest IG
3. recur over the subsets given by splitting at the feature and threshold with the highest IG

progress

progress

conclusion

- intuition and heuristics needed for each algorithm
- choice of algorithm depends on features of interest
- future work:
  - alternative distance metrics
  - normalize input data
  - parallelize algorithms

references


Yann LeCun and Corinna Cortes.
MNIST handwritten digit database.
2010.


Yale face database b.

thanks!