Branching Out Carefully: A Cautious Take on Random Forests

Machine Learning Final Project

David Feldman, Jarrett Huddleston, Tyler Skow

Motivation

Motivation

Cat or dog?

Motivation

Hmm…
I'm 51% sure that's a cat. Therefore, I will say with **certainty**, its a **cat**.

Motivation

Maybe that's not what we want…

Motivation

Guiding Questions

- How do you account for uncertainty during decision-making?
- Can we reward a classifier for exercising cautiousness?
- Can abstention improve accuracy in certain contexts? How can this be measured?

Research goals and contributions

- Demonstrate how cautious classification can enhance Random Forests when erroneous classification is especially costly
- Extend existing methods to cautious classification with conformal prediction and fuzzy Random Forests
- Develop tailored metric to assess tradeoffs between abstention and accuracy of cautious models

Methods

- Compared Baseline RF and several cautious RF implementations
  - **Baseline RF**:
    - Used Sci-kit Learn, grid search CV to pick optimal hyperparameters
  - **Naive Cautious Classifier**:
    - Built off Baseline RF
    - Discarded predictions where confidence below user-defined threshold
    - Also used grid search CV to pick optimal hyperparameters
  - **Naive Cautious Classifier with Class Specific Thresholds**:
    - Extends by adding class-specific thresholds (fit via grid search CV)
    - Prevents issues from fitting one threshold to potentially unbalanced data

Methods

- **Cautious Weighted RF**:
  - Fits basebine RF, calculates interval based probability estimates, uses them to calculate measures of belief and plausibility
  - If observation belief>0.5: 1. If observation plausibility<0.5: 0. If neither: undetermined
  - Also used grid search CV to pick optimal hyperparameters
- **Conformal Prediction**:
  - Generates prediction sets with predefined coverage guarantee, labels observations undetermined where intervals overlap
  - Also supports nonconformity scores

Methods

- **Fuzzy RF**:
  - Utilizes ability of Fuzzy DT to handle data uncertainty and bagging efficacy to create more accurate classifier where data and labels may not be clearly lineated
  - Popular technique for data evaluation where data may not be present

Evaluation Metrics

- Need metrics that reward/penalize appropriately based on model's accuracy, abstention rate
- **Single-set accuracy**: accuracy on set of points classifier makes decision on
- **Determinacy**: proportion of observations where model makes determinate prediction
- **Abstention**: 100-Determinacy (proportion of observations model labels as indeterminate)
- **U65**: Discounted accuracy measure accounting for accuracy and determinacy
- For Baseline RF: used accuracy score

Data

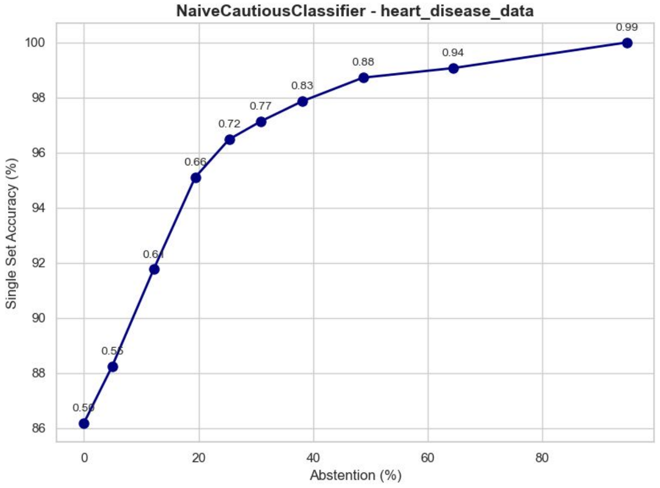
- Ran models on 4 different datasets
  - German Credit Data
    - Predicts whether person good or bad creditor
  - Breast Cancer Data
    - Predict whether person has cancer
  - Heart Disease Data
    - Predicts whether person has heart disease
  - COMPAS Recidivism
    - Predicts two-year recidivism
- Each dataset split into training, testing, calibration sets (See report for full data pre-processing steps and missing data analysis)

Results - Single Set Accuracy

- The conformal predictor (CP) beats the CWRF by 7.88%
- The naive cautious classifier (NCC) beats the CWRF by 8.9%
- NP, NCC, & CWRF match or beat other models
- Seem like the best models, except?

Results - Determinacy

- Standard RF and Fuzzy RF always 100% determinacy
- Conformal predictor and naive cautious classifier never at 100% determinacy
  - Conformal predictor: 41.92%
  - Naive cautious classifier: 65.24%
- WCRF has 100% determinacy in ¾ of the datasets
  - Still 78.91% in the fourth



Results - U65 Score

- We need to consider the U65 Score
  - WCRF more closely matches or beats CP and NCC
    - In one case by over 20% with the breast cancer dataset
  - While matching or outperforming non-cautious classifiers

Classifier	Dataset	U65 Score	Single Set Accuracy	Determinacy	Precise Accuracy	Abstention
RandomForest	german_credit_data	74.0	74.0	100.0	74.0	0.0
ConformalPredictor	german_credit_data	74.52	87.95	41.5	74.0	58.0
CWRF	german_credit_data	76.0	76.0	100.0	78.0	0.0
NaiveCautiousClassifier	german_credit_data	77.25	87.27	55.0	74.0	45.0
CSTCautiousClassifier	german_credit_data	75.48	86.6	48.5	74.0	51.5
FuzzyRandomForest	german_credit_data	70.0	70.0	100.0	70.0	0.0

Future Work

- Multiclass analysis
- Spend more time tuning models
- Incorporate MNAR data
- Consider more datasets

Questions?