Improved Algorithms for Confidence-Rated Prediction with Error Guarantees

Kamalika Chaudhuri and Chicheng Zhang
kamalika@cs.ucsd.edu, chichengzhang@ucsd.edu

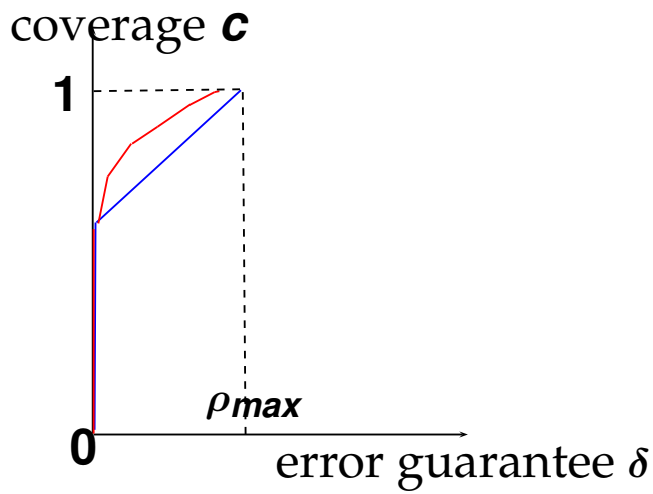
Dec. 10th, 2013

Confidence-Rated Prediction: Motivation

- In some applications, it is better to say *Don't Know* than make a mistake.
- Example: medical diagnosis software
- Goal: Design a predictor which can say *Don't Know*
- Predictor can be to identify *easy* data to classify.

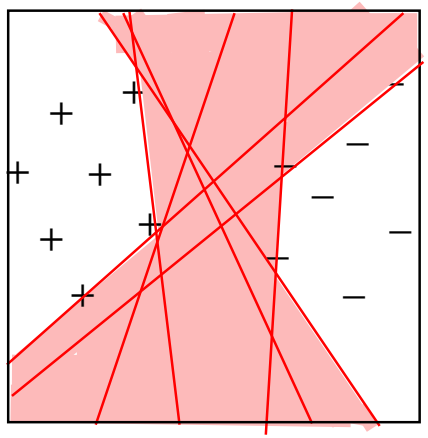
Two performance metrics

- Error: fraction of data on which the predictor outputs the wrong label.
  Coverage: fraction of data on which predictor outputs a label $\{$-1,+1$\}$;
- Error-coverage tradeoff!
- Goal:
  - given error guarantee δ, maximize coverage $\boldsymbol{c}$
  - An algorithm(strategy): take δ as a parameter $\Rightarrow$ a δ - $\boldsymbol{c}$ curve



Algorithm

- Realizable case
- A Candidate Strategy[EYW10]
  - 1. Compute a version space $\boldsymbol{V}$.
  - 2. If all $\boldsymbol{h} \in \boldsymbol{V}$ agree on the label of $\boldsymbol{x}$, predict its label by $\boldsymbol{h}(\boldsymbol{x})$ using an arbitrary $\boldsymbol{h} \in \boldsymbol{V}$
  - 3. Else predict *Don't Know* .



- Randomize on examples in disagreement region to get a curve of δ - $\boldsymbol{c}$ tradeoff
- Is it optimal? No, at least in *transductive* setting
- Observation: optimal prediction can be formulated as a linear program in transductive setting
- maximize coverage, subject to the error upper bound estimated by $\boldsymbol{V} \leq \delta$.
- predict using solution of the linear program
- Can be extended to agnostic setting
- See our poster for details & experimental results