What to Expect of Classifiers?

Reasoning about Logistic Regression with Missing Features

Pasha Khosravi Yitao Liang Yoojung Choi Guy Van den Broeck
(pashak, yliang, yjchoi, guyvdb)@cs.ucla.edu

Motivation

Classifiers are generally not able to make predictions in presence of uncertainty over input features \( X \)

\[ \Rightarrow \text{e.g., with missing values!} \]

The probabilistic way to deal with this, is to **compute the expected predictions** of a classifier given a feature distribution. That is, we want to classify a partial sample \( X \).

How hard is computing expectations?

Surprisingly computing expectations is hard for even simple classifiers and distributions:

- \( \mathcal{F} \) is a nontrivial classifier and \( P \) is uniform \( \Rightarrow \#P\text{-Hard} \)
- \( \mathcal{F} \) is a single-feature classifier and \( P \) is an arbitrary PGM \( \Rightarrow \#P\text{-Hard} \)
- \( \mathcal{F} \) is a logistic regressor and \( P \) Naive Bayes \( \Rightarrow \text{we prove it to be NP-Hard!} \)

Conformant Learning

We say \( P(X, C) \) **conforms** with \( \mathcal{F} : \mathcal{X} \rightarrow [0, 1] \) if their classifications agree: \( P(c \mid x) = \mathcal{F}(x) \) for all \( x \).

Conformant learning finds the generative model \( P \theta(X, C) \) which conforms to a classifier \( \mathcal{F}(x) \) and maximises the feature likelihood:

\[
\arg\max_{\theta} \prod_{d \in [x] \in D} \sum_{c} P_\theta(x, c) \quad \text{s.t.} \forall x : P_\theta(c \mid x) = \mathcal{F}(x)
\]

Naive Conformant Learning (NaCL) employs a Naive Bayes model for \( P \) and a Logistic Regressor for \( \mathcal{F} \)

\[ \Rightarrow \text{efficiently solvable as geometric programming} \]

References