UCLA Computer Science

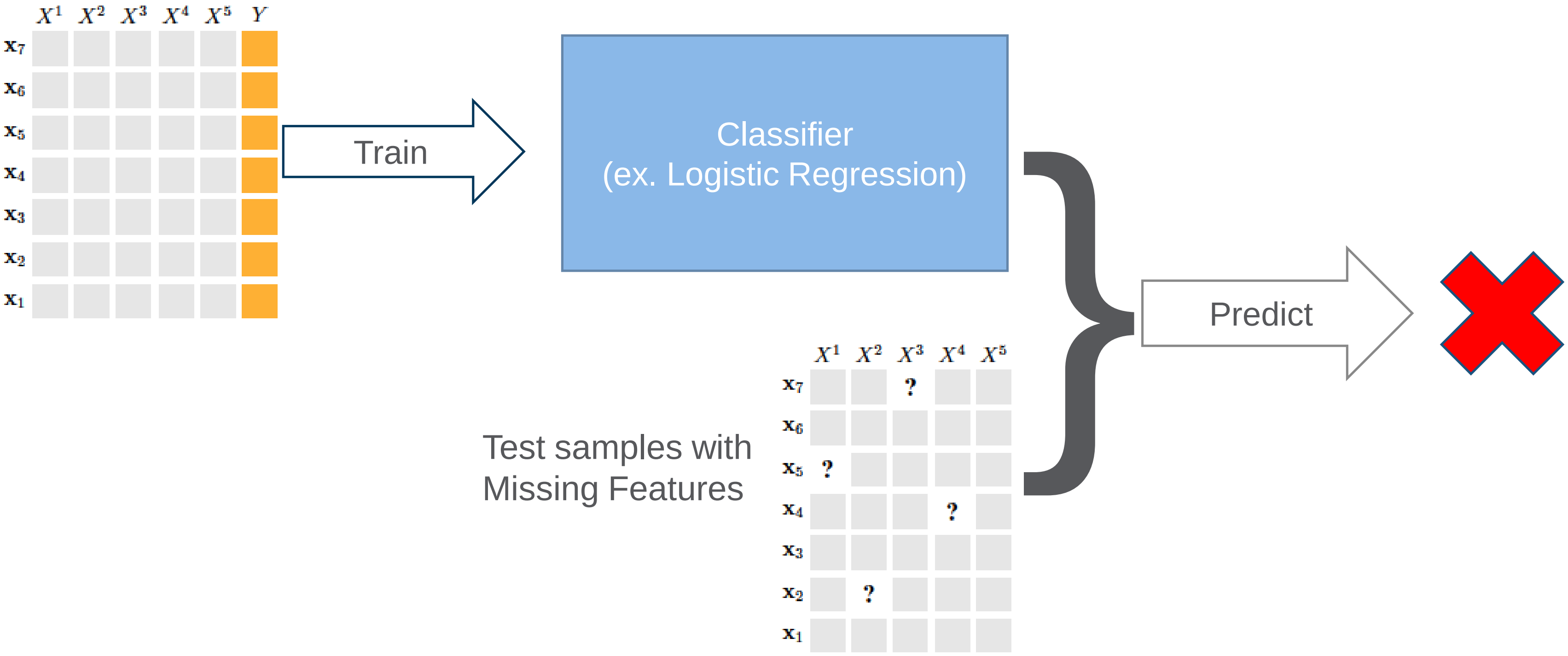
STAR AI RESEARCH LAB UCLA

What to Expect of Classifiers?

Reasoning about Logistic Regression with Missing Features

Pasha Khosravi, Yitao Liang, YooJung Choi and Guy Van den Broeck.
Computer Science Department, UCLA

Motivation



X^1 X^2 X^3 X^4 X^5 Y
$\mathbf{x}_7$ $\mathbf{x}_6$ $\mathbf{x}_5$ $\mathbf{x}_4$ $\mathbf{x}_3$ $\mathbf{x}_2$ $\mathbf{x}_1$
Train
Classifier
(ex. Logistic Regression)
Predict
Test samples with
Missing Features
X^1 X^2 X^3 X^4 X^5
$\mathbf{x}_7$ $\mathbf{x}_6$ $\mathbf{x}_5$ $\mathbf{x}_4$ $\mathbf{x}_3$ $\mathbf{x}_2$ $\mathbf{x}_1$
?
?
?
?

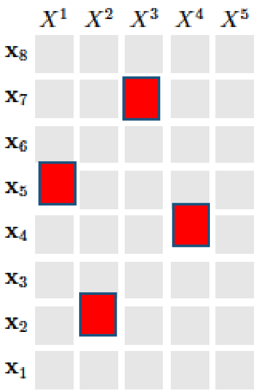
Common Approaches

- Common approach is to fill out the missing features, i.e. doing imputation.

- They make unrealistic assumptions (mean, median, etc).

- More sophisticated methods such as MICE don’t scale to bigger problems (also have assumptions).

- We want a more principled way of dealing with this while staying efficient.



Generative vs Discriminative Models

	Discriminative Models (ex. Logistic Regression)	**Generative Models (ex. Naïve Bayes)**
	$P(C \mid X)$	$P(C, X)$
Missing Features	😕	🙂
Classification Accuracy	🙂	😕

Expected Predication

- How can we leverage both discriminative and generative models?

- “Expected Prediction” is a principled way to reason about outcome of a classifier, $F(X)$, w.r.t. a feature distribution $P(X)$.

$$E_{\mathcal{F},P}(\mathbf{y}) = \mathop{\mathbb{E}}_{\mathbf{m}\sim P(\mathbf{M}|\mathbf{y})} \left[\mathcal{F}(\mathbf{ym})\right]$$

M: Missing features
y: Observed Features

Expected Predication Intuition

- **Imputation Techniques**: Replace the missing-ness uncertainty with *one* or *multiple* possible inputs, and evaluate the models.

- **Expected Prediction**: Considers *all possible inputs* and reason about expected behavior of the classifier.

$$E_{\mathcal{F},P}(\mathbf{y}) = \sum_{\mathbf{m}} P(\mathbf{m} \mid \mathbf{y}) \cdot \mathcal{F}(\mathbf{ym}) = \mathop{\mathbb{E}}_{\mathbf{m} \sim P(\mathbf{M}|\mathbf{y})} [\mathcal{F}(\mathbf{ym})]$$

Hardness of Taking Expectations

- How can we compute the expected prediction?

- In general, it is intractable for arbitrary pairs of discriminative and generative models. ☹

- Even when F is Logistic Regression and P is Naïve Bayes, the task is NP-Hard. ☹

Conformant learning

Given a discriminative classifier and a dataset, learn a generative model that

1. *Conforms* to the classifier.

2. Maximizes the likelihood of joint feature distribution P(X)

No missing features → Same quality of classification ☺
Has missing features → No problem, do inference ☺

Naïve Conformant Learning (NaCL)

We focus on of Conformant Learning involving Logistic Regression and Naïve Bayes

- Given a NB model there is unique LR model that conform to it
- Given a LR model there is many NB models that conform to it

Naïve Conformant Learning (NaCL)

- We showed that we can write the Naïve Conformant Learning Optimization task as a *Geometric Program.*

- ***Geometric Programs*** are a special type of constraint optimization problems that have an exact and efficient algorithm to optimize, and modern GP solvers can handle large problems.

- For NaCL, we have $O(nk)$ number of constraints. n is the number of features, and k is the number of classes.

Naïve Conformant Learning (NaCL)

Logistic Regression Weights

NaCL

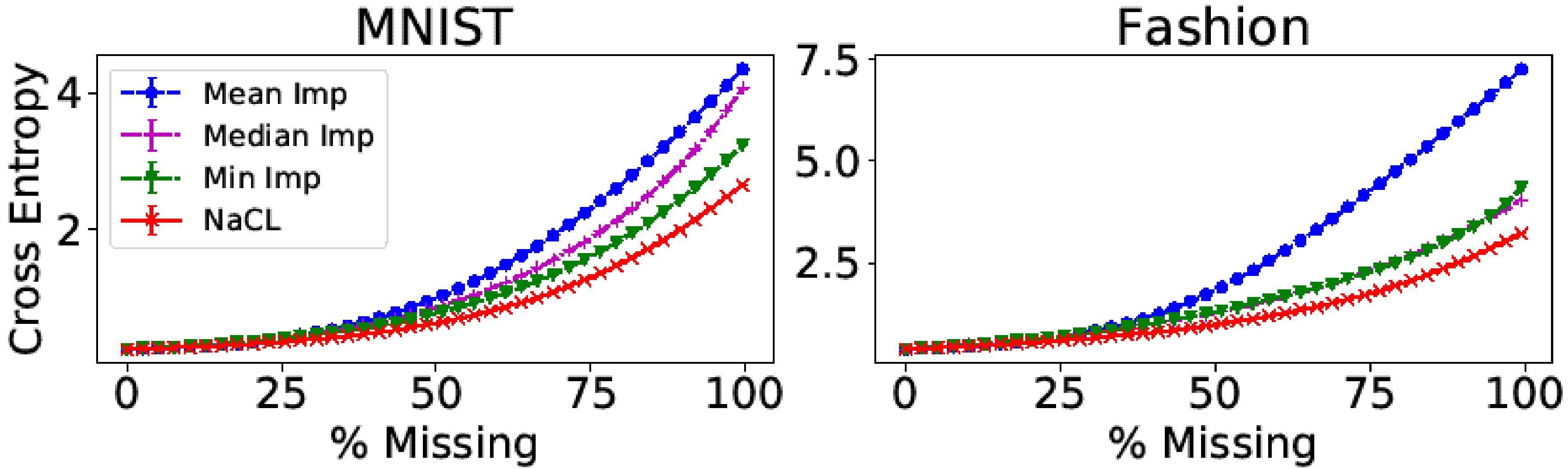
"Best" Conforming Naïve Bayes

GitHub: github.com/UCLA-StarAI/NaCL

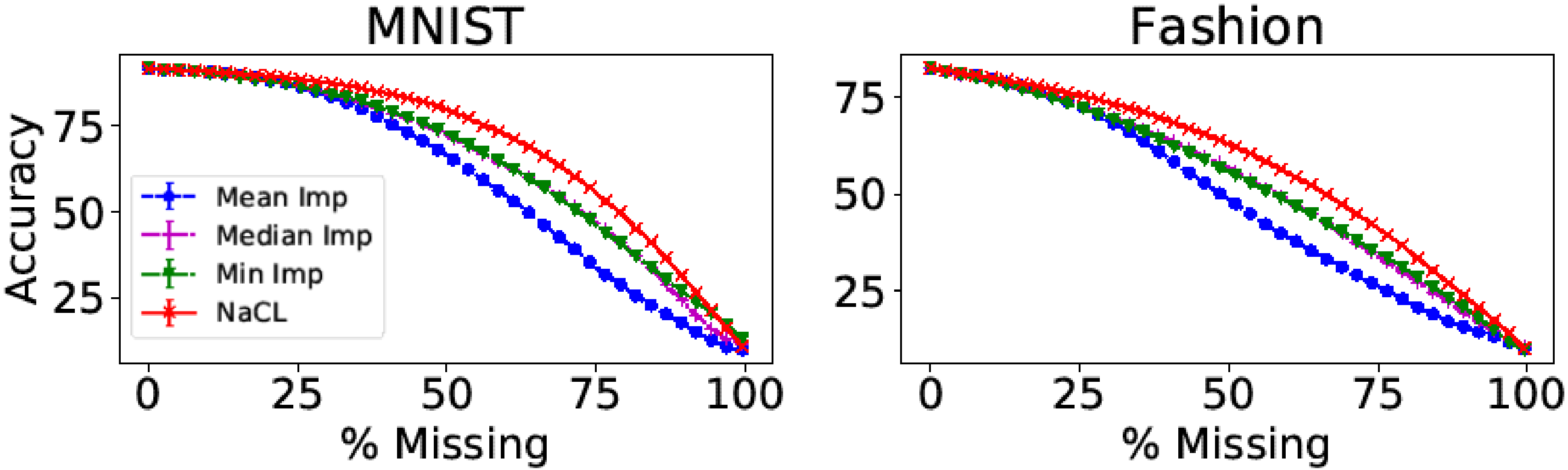
Experiments: Fidelity to Original Classifier

Using Cross Entropy to compare

- probabilities of the original classifier vs probabilities of NaCL's learned model



Experiments: Classification Accuracy



Other Applications

We saw *Expected Prediction* is very effective with handling missing features.

What else can we do?

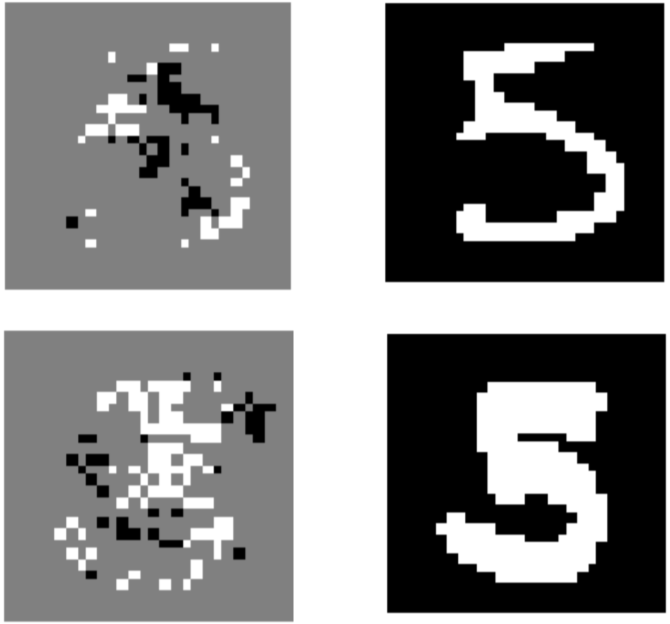
- Explanations
- Feature Selection
- Fairness

Local Explanations using Missing-ness

Goal: To explain an instance of classification

- *Support Features*:
  Making them missing → probability goes **down**

- *Opposing Features*:
  Making them missing → probability goes **up**



Sufficient Explanations

Remove maximum number of supporting features until expected classification is about to change, then show the remaining support features.

Conclusion

- Expected Prediction is an effective tool for several applications such as missing data, generating explanations

- We introduced NaCL, an efficient algorithm, to convert a Logistic Regression model to a conforming Naïve Bayes model.

- Future work would be looking at more expressive pair of models, and potentially choose models that make the expected prediction tractable.

Thank You

What to Expect of Classifiers? Reasoning about Logistic Regression with Missing Features

GitHub: github.com/UCLA-StarAI/NaCL