LOREM IPSUM
DOLOR SIT AMET
Subtitle

Thursday 22nd March, 2018

Prénom Nom

LIP6

MLIA Machine Learning &
Deep Learning for
Information Access

Noise filtering in UPMC Food-101

Principle

The problem

- UPMC Food-101 has been crawled from Google Images
- It contains a certain amount of noise

The idea

- Creating bags from images of 1 class (e.g. pizza)
- Creating bags for "rest" class
- Defining the expected level of noise in the pizza bags
- Using *Learning with Label Proportions* models to detect noise

Experimentations

SCIENCES
SORBONNE
UNIVERSITÉ

Protocol

- **Dataset:** $(x_i, y_i, y_i^*) \in \mathbb{R}^p \times \{-1, 1\} \times \{-1, 1\}, \quad i = 1..n$
  - x_i features, y_i noisy label, y_i^* true label (not available for training)
- **Create bags:** Create bags b_j of 30 points having the same y_i and give them a proportion of positive points p_j
- **Training:** Train the SyMIL model on bags and train an SVM on (b_j, p_j) / (x_i, y_i)
- **Evaluation:** Use the decision frontiers of SyMIL / SVM models to reclassify each x_is by predicting $\hat{y}_i^*$, and compare with y_i^*

Some papers for noisy images datasets

"Auxiliary Image Regularization for Deep CNNs with Noisy Labels" Azadi et al. 2015

- Define a regularized loss for training the CNN
- Can be seen as looking for the label of similar images for regularization
- Results slightly better than Sukhbaatar model

References

References I

Azadi, Samaneh et al. (2015). "Auxiliary Image Regularization for Deep CNNs with Noisy Labels". In: arXiv: 1511.07069.