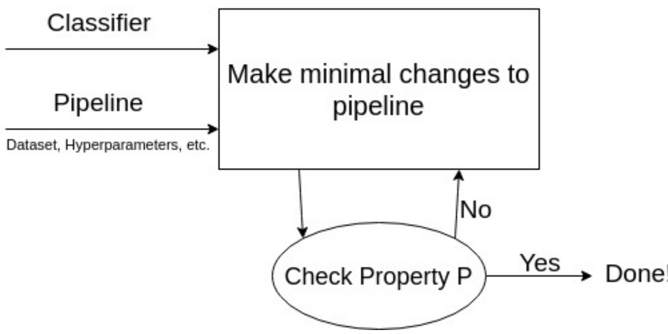
CS-799: Notes

Goutham Ramakrishnan
Guide: Prof. Aws Albarghouthi

Broad goal:
Given a machine learning pipeline with a dataset and classifier, find a minimal set of changes which need to be made to the pipeline for it to satisfy a given property P.



Ideas:

- Use clustering (K-Means? Hierarchical? Semi-supervised?)
- Make changes to dataset iteratively
- Changes made must be interpretable
- Use training data debugging using trusted items as Property P to be starting point for exploring approaches

Discussion with Xuezhou:

- Using clustering as starting point for DUTI
- Active Debugging: Prof. Po-Ling's student working on it. Requires periodic expert feedback.
- For interpretability, find bugs first and then learn **rule list**
- Debugging the machine learning pipeline: Talk by Prof. Jerry Zhu (http://pages.cs.wisc.edu/ jerryzhu/adversarial/pub/debugML.pdf)
- Difficult to test approach due to lack of good datasets for this kind of task

Thoughts on clustering:

- Advantage of hierarchical clustering over K-means: The output hierarchy is more informative than output of K-means, number of clusters need not be pre-specified, can use dendogram to choose K. Tradeoff: Speed and efficiency
- Review on Semi-supervised clustering methods: https://arxiv.org/pdf/1307.0252.pdf
- Constrained clustering looks promising, can enforce cannot-link constraints on differently labeled trusted items
- Need to read up more on semi-supervised hierarchical clustering

Formulation:
Training set: $(X, Y) = \{(x_i, y_i)\}_{1:n}$
Trusted set: $(X, Y) = \{(\widetilde{x}_i, \widetilde{y}_i\}_{1:m}$
Number of clusters: K, Set of clusters $S = \{S_1, \ldots, S_K\}$

K-Means:

$$\underset{S}{\operatorname{argmax}} \sum_{k=1}^{K} \sum_{x \in S_i} \| x - \mu_k \|_2^2 = \underset{S}{\operatorname{argmin}} \sum_{k=1}^{K} |S_k| Var(S_i)$$

To minimize variance, add:

$$\sum_{k=1}^{K} Entropy(y_i | x_i \in C_k)$$

To ensure trusted items with different labels are not in the same cluster, add:

$$\sum_{\widetilde{x_i}, \widetilde{x_j} : \widetilde{y_i} \neq \widetilde{y_j}} Cost(\widetilde{x_i}, \widetilde{x_j}) \mathbb{1}(C(\widetilde{x_i}) = C(\widetilde{x_j}))$$

$C(\widetilde{x_i})$ is the cluster to which $\widetilde{x_i}$ belongs.