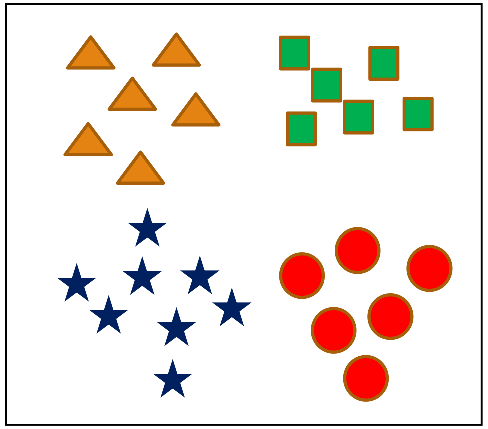
Multiclass Classification and Weak Supervision

Multiclass Classification

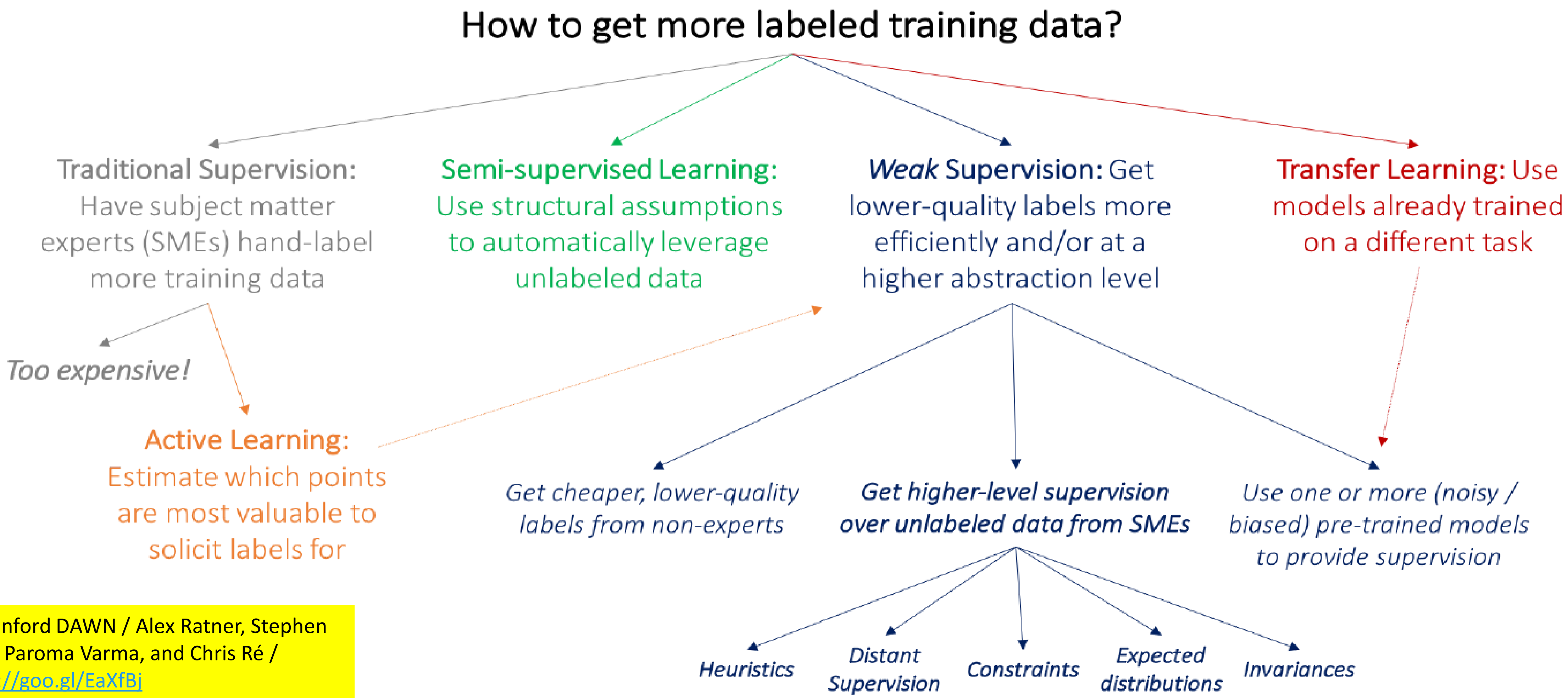


- Classification involving more than two classes (i.e., > 2 Classes)
- Methodology
  - Reducing the multi-class problem into multiple binary problems
- Method 1. **One-vs.-rest** (or **one-vs.-all**)
  - Given *m* classes, train *m* classifiers: One for each class
  - Classifier j: Treat tuples in class j as *positive* & **all the rest** as *negative*
  - To classify a tuple **X**, the set of classifiers vote as an ensemble
- Method 2. **One-vs.-one** (or **all-vs.-all**): Learn a classifier for each pair of classes
  - Given *m* classes, construct $m(m-1)/2$ binary classifiers
  - A classifier is trained using tuples of the two classes
  - To classify **X**, each classifier votes: **X** is assigned to the class with maximal vote
- Comparison: One-vs.-one tends to perform better than one-vs.-rest
- Many new algorithms have been developed to go beyond binary classifier method

Weak Supervision: A New Programming Paradigm for Machine Learning

- Overcome the training data bottleneck
  - Leverage higher-level and/or noisier input from experts
- Exploring *weak label distributions* provided more cheaply and efficiently by
  - Higher-level, less precise supervision (e.g., heuristic rules, expected label distributions)
  - Cheaper, lower-quality supervision (e.g., crowdsourcing)
  - Existing resources (e.g., knowledge bases, pre-trained models)
- These weak label distributions could take many forms
  - Weak labels from crowd workers, output of heuristic rules, or the result of distant supervision (from KBs), or the output of other classifiers, etc.
  - Constraints and invariances (e.g., from physics, logic, or other experts)
  - Probability distributions (e.g., from weak or biased classifiers, user-provided labels, feature expectations, or measurements)

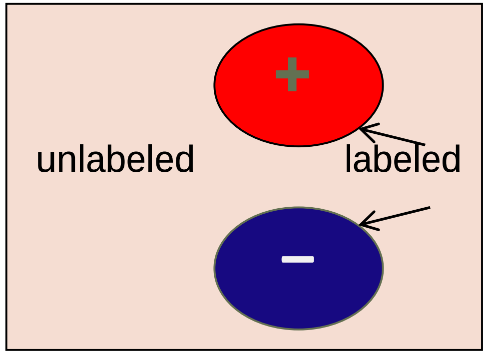
Relationships Among Different Kinds of Supervisions



Many areas of machine learning are motivated by the bottleneck of labeled training data, but are divided at a high-level by what information they leverage instead.

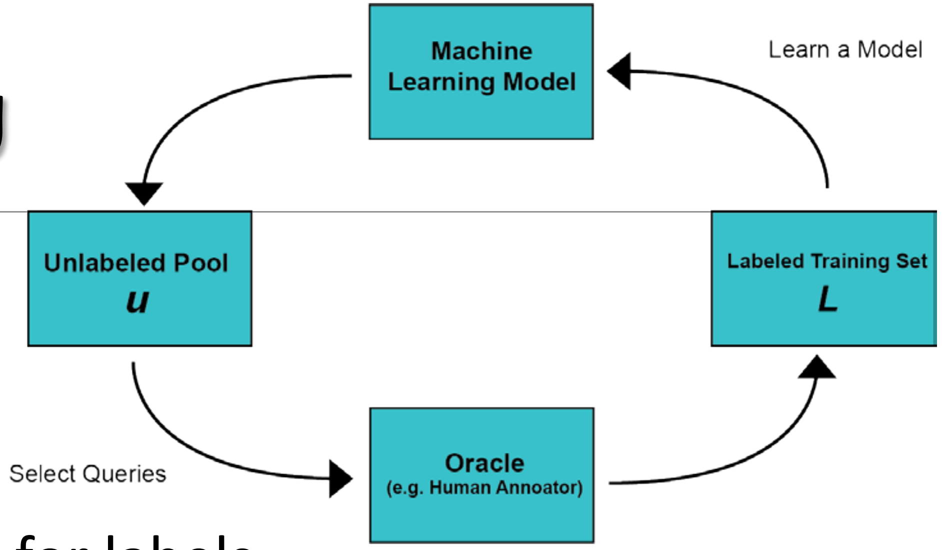
Semi-Supervised Classification

- Semi-supervised: Uses labeled and unlabeled data to build a classifier
- Self-training
  - Build a classifier using the labeled data
  - Use it to label the unlabeled data, and those with the most confident label prediction are added to the set of labeled data
  - Repeat the above process
  - Adv.: Easy to understand; Disadv.: May reinforce errors



- Co-training: Use two or more classifiers to teach each other
  - Each learner uses a mutually independent set of features of each tuple to train a good classifier, say f_1 and f_2
  - Then f_1 and f_2 are used to predict the class label for unlabeled data X
  - Teach each other: The tuple having the most confident prediction from f_1 is added to the set of labeled data for f_2 & vice versa
- Other methods include joint probability distribution of features and labels

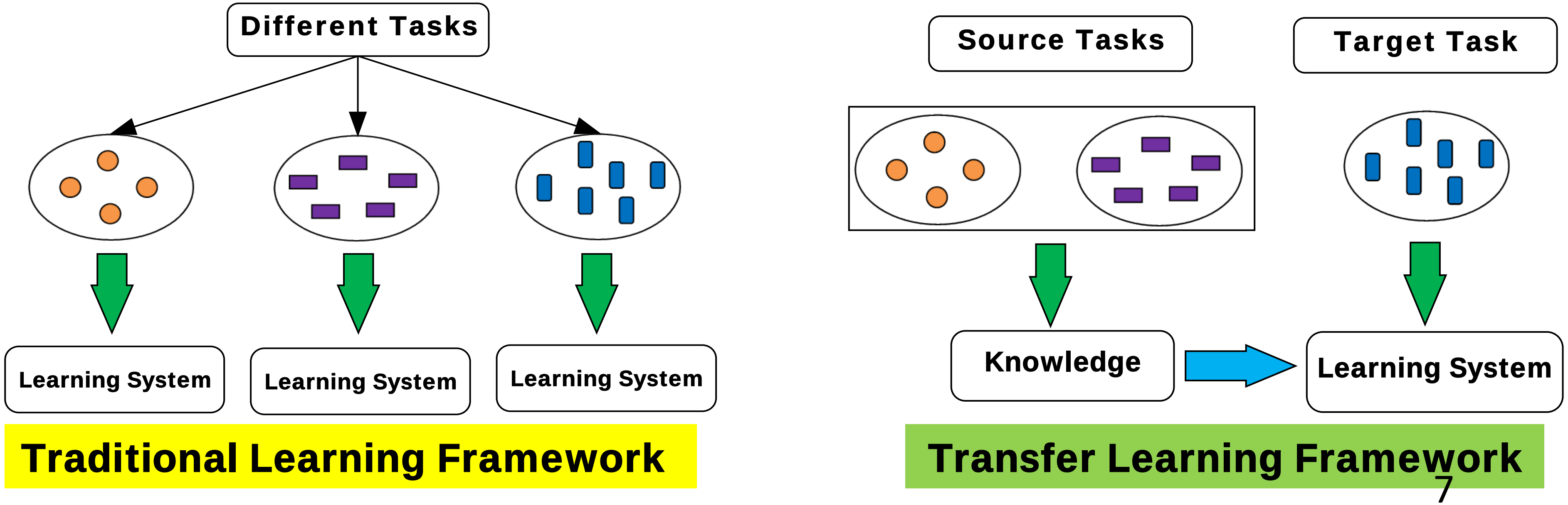
Active Learning



- A special case of semi-supervised learning
  - Unlabeled data: Abundant
  - Class labels are expensive to obtain
- Active learner: Interactively query teachers (oracle) for labels
- Pool-based approach: Uses a pool of unlabeled data
  - L: A small subset of D is labeled; U: A pool of unlabeled data in D
  - Use a query function to carefully select one or more tuples from U and request labels from an oracle (a human annotator)
  - The newly labeled samples are added to L, and learn a model
  - Goal: **Achieve high accuracy using as few labeled data as possible**
- Evaluated using *learning curves*: Accuracy as a function of the number of instances queried (# of tuples to be queried should be small)
- A lot of algorithms have been developed for active learning

Transfer Learning: Conceptual Framework

- Transfer learning: Extract knowledge from one or more source tasks (e.g., recognizing cars) and apply the knowledge to a target task (e.g., recognizing trucks)
- Traditional learning: Build a new classifier for each new task
- Transfer learning: Build new classifier by applying existing knowledge learned from source tasks
- Many algorithms are developed, applied to text classification, spam filtering, etc.



7

Summary

Summary

- Model Evaluation and Selection
- Techniques to Improve Classification Accuracy: Ensemble Methods
- Multiclass Classification and Weak Supervision

Recommended Readings

- Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123-140. Retrieved from https://link.springer.com/article/10.1023/A:1018054314350
- Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. London, UK: Chapman & Hall/CRC.
- Freund, Y. & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *JCSS, 55*(1), 119-139. Retrieved from https://www.sciencedirect.com/science/article/pii/S002200009791504X
- Gao, J., Fan, W., & Han, J. (2007). A general framework for mining concept-drifting data streams with skewed distributions. *Proc. of SDM*. DOI: 10.1137/1.9781611972771.1. Retrieved from http://epubs.siam.org/doi/abs/10.1137/1.9781611972771.1
- Grossman, R., Seni, G., Elder, J., Agarwal, N., & Liu, H. (2010). *Ensemble methods in data mining: Improving accuracy through combining predictions*. San Rafael, CA: Morgan & Claypool.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI, 2*, 1137-1143. Retrieved from https://dl.acm.org/citation.cfm?id=1643047
- Pan, S. J. & Yang, Q. (2010). A Survey on transfer learning. *IEEE Trans. on Knowledge and Data Eng.* Retrieved from http://ieeexplore.ieee.org/document/5288526/
- Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *Int. Journal of Pattern Recognition and Artificial Intelligence, 23*(4), 687. Retrieved from http://www.worldscientific.com/doi/abs/10.1142/S0218001409007326
- Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms*. Boca Raton, FL: CRC Press.

References

- Stanford DAWN. (2017). *Labeled training data graphic* [Online image]. Retrieved from https://goo.gl/EaXfBj