BITS F464: Machine Learning

30 Active Learning, Metric Learning

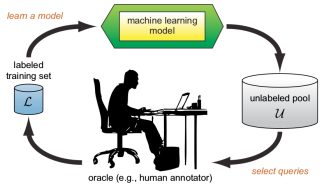
Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

April 07, 2021 ONLINE (Campus @ BITS-Pilani Jan-May 2021)

http://ktiwari.in/ml

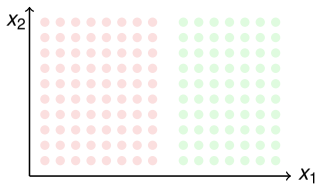
Active Learning

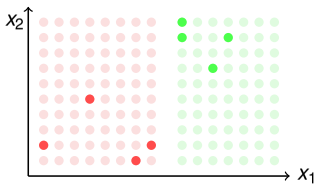
Active Learning is a **semi-supervised** approach where learner asks questions of his interest (choose the data from which to learn)



- Perform better with less training. (why to focus on less training?)
- Obtaining labeled data is difficult, time-consuming, and expensive
- Consider labeling speech utterances, one minute of speech may take seven hours to annotate

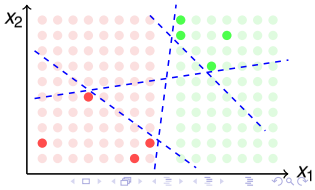
Instance space and Training data



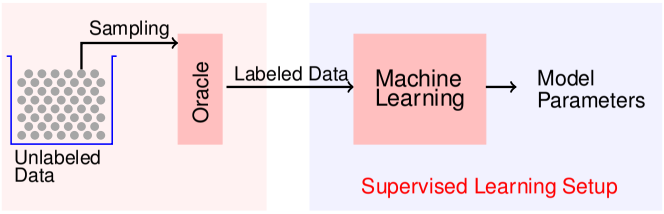


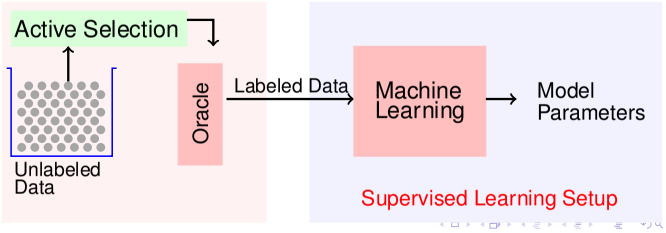
- Points in training data are few
- Many consistent boundaries
- All data point may NOT equally contribute to the learning
- Region of disagreement (RoD)

Test if a point is in RoD is difficult



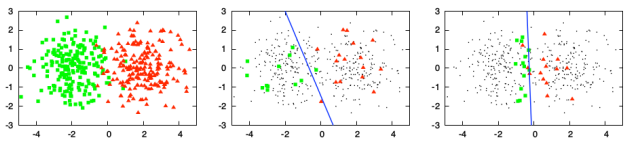
Active Learning: Add smart data selection mechanism





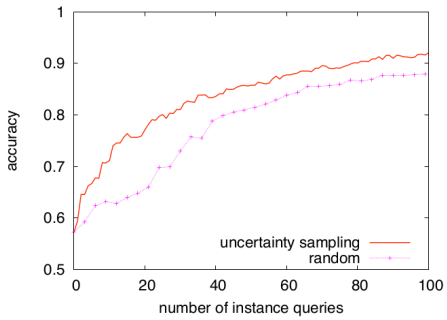
Active Learning

- Questions are
  - When to query
  - What to query
- **Uncertainty sampling:** selects instance about which the model is least certain (point in RoD)



Example: Data, logistic regression with 30 randomly drawn labels; accuracy: 70%, with 30 actively queried instances accuracy: 90%).

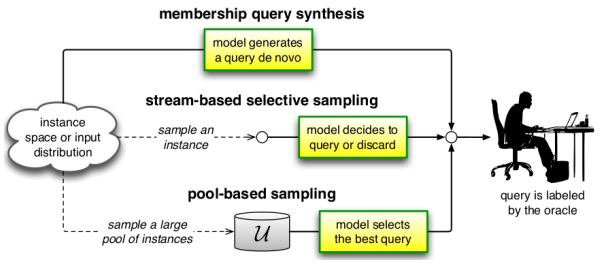
Active Learning



- For 1-D linearly separable data, binary search is useful
- With n training samples, error reduces by $1/n$ and 2^{-n} for passive and active learning respectively.

Active Learning Scenarios

There are several different scenarios in which the learner may be able to ask queries



Query Strategy

- **Uncertainty Sampling:**

$$x^*_{LC} = \underset{x}{\operatorname{argmax}}\, 1 - P_\theta(\hat{y}|x)$$

- **Query-By-Committee:** a committee of models $\{\theta^{(1)}, \theta^{(2)}, ..., \theta^{(C)}\}$ is maintained. Instance having most disagree are queried.
- **Expected Model Change:** selecting the instance that would impart the greatest change to the current model.
- **Expected Error Reduction:** selecting the instance that reduce generalization error the most
- **Variance Reduction:** learning objective is to minimize standard error
- **Density-Weighted Methods:** less prone to querying outliers

Metric Learning

- Learning in many applications depends upon **proximity measures**
- Distance measure is used to find dissimilarity between pattern representations. More similar patterns should be near
- Metric measure has following three properties
  1. *Positive reflexivity* $d(x, x) = 0$
  2. *Symmetry* $d(x, y) = d(y, x)$
  3. *Triangular Inequality* $d(x, y) \leq d(y, z) + d(z, y)$
- **Minkowski metric:**

$$d^m(X, Y) = \Big(\sum_{k=1}^{d} |x_k - y_k|^m\Big)^{\frac{1}{m}}$$

- Manhattan distance or L_1 uses $m = 1$, Euclidean or L_2 has $m = 2$
- L_0 is number of non-zero elements and L_∞ is maximum

More approached

- **Weighted Distance**

$$d^m(X, Y) = \Big(\sum_{k=1}^{d} w_k |x_k - y_k|^m\Big)^{\frac{1}{m}}$$

- When w_k is the variance σ^2 across kth attribute then it is Mahalanobis distance
- **Example of a non Metric measure:**
  1. k-median
  2. Cosine

$$S(X, Y) = \frac{X^t Y}{||X||\ ||Y||}$$

  3. KL-Distance

$$KL(p, q) = \sum_i p_i \log_2 \left(\frac{p_i}{q_i}\right)$$

Thank You!

Thank you very much for your attention!

Queries ?

(Reference[1])

[1] [1] Settles, Burr (2010), "Active Learning Literature Survey" (PDF), Computer Sciences Technical Report 1648. University of WisconsinMadison, retrieved 2014-11-18 [2] Papers stated in slides