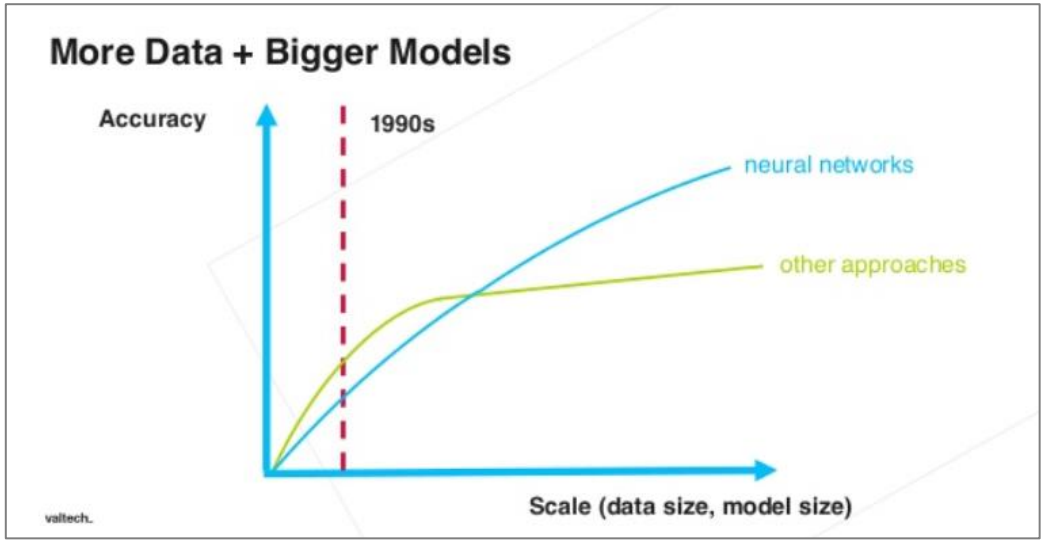
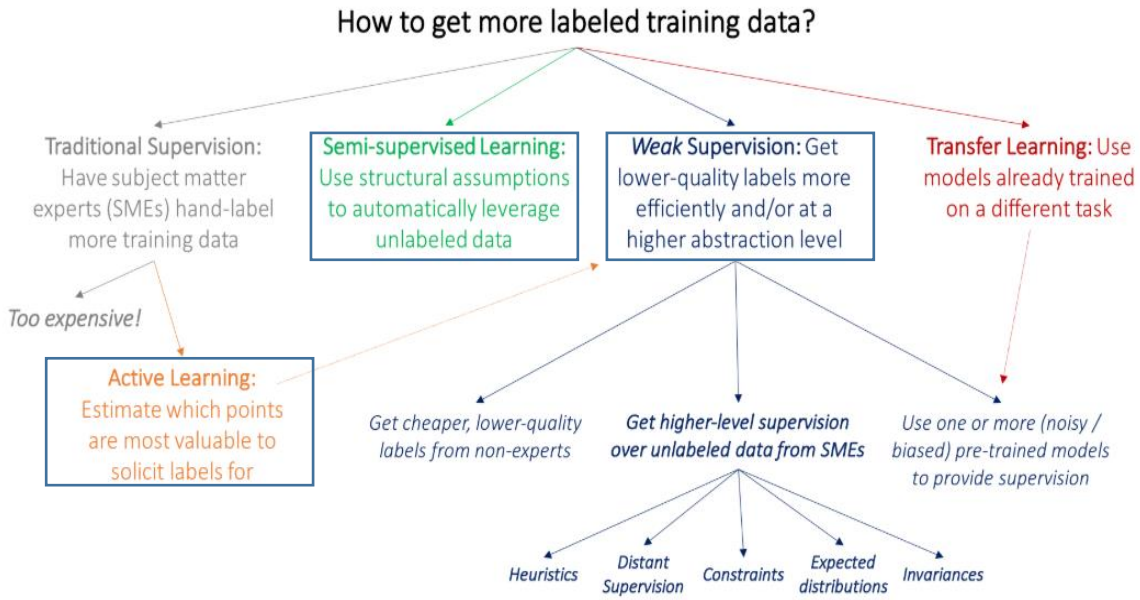
Brief Introduction to Weakly Supervision

The Training Data Bottleneck



- Deep learning models are massively complex – millions of free parameters – and thus require more labeled training data.
- Cost and Inflexibility of hand-labeling such training sets is the key bottleneck to actually deploying machine learning.
- The reason why we use weak supervision.

What is Weakly Supervision

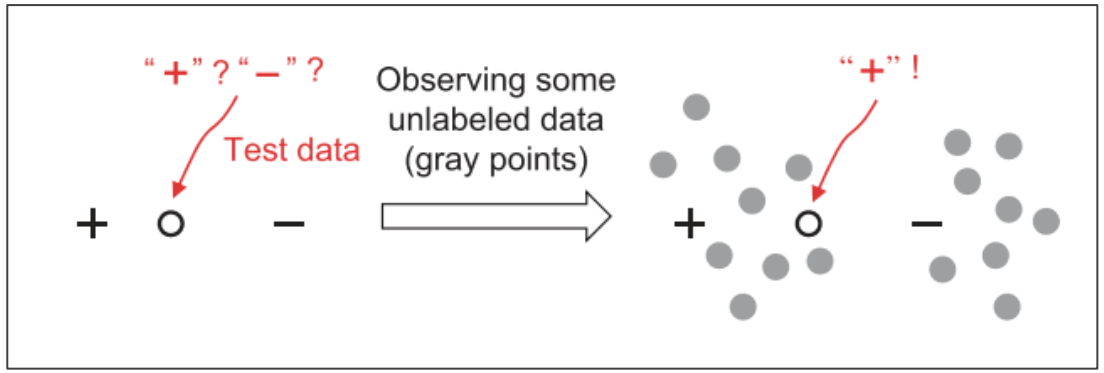


Weak supervision : How to get lower quality labels more efficiently and at a higher abstraction level

Strong supervision : information like fully ground-truth labels

Why it works

Ex)



$$f(\boldsymbol{x}|\Theta) = \sum_{j=1}^{n} \alpha_j f(\boldsymbol{x}|\theta_j),$$

Assume data com from Gaussian mixture model with n mixtures,

$$h(\boldsymbol{x}) = \arg\max_{c\in\{Y,N\}} \sum_{j=1}^{n} P(y_i = c|g_i = j, \boldsymbol{x}_i) \times P(g_i = j|\boldsymbol{x}_i),$$

According to maximum a posteriori criterion,

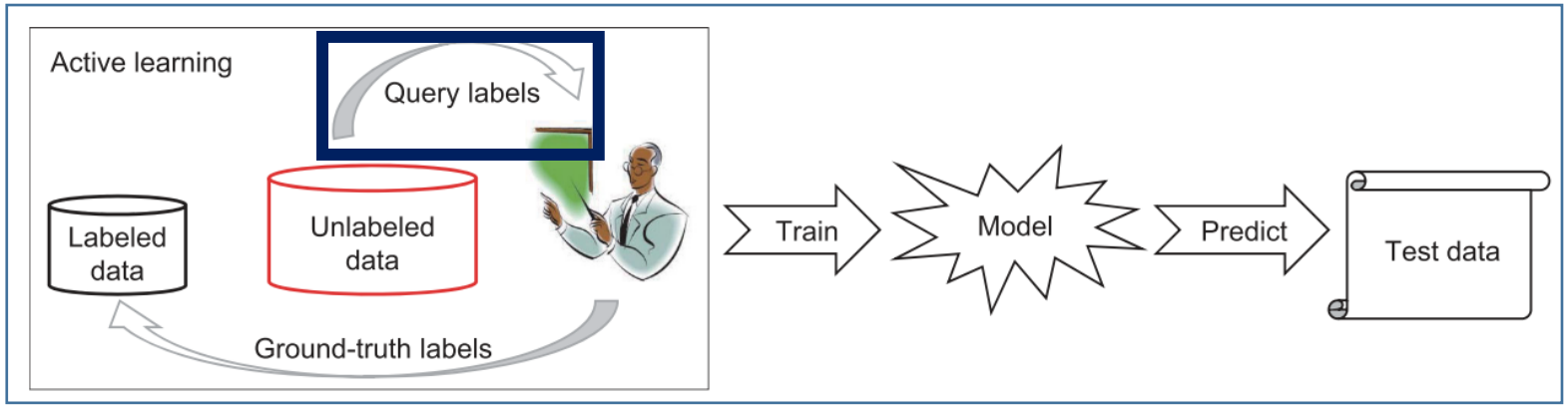
$$P(g_i = j|\boldsymbol{x}_i) = \frac{\alpha_j f(\boldsymbol{x}_i|\theta_j)}{\sum_{k=1}^{n} \alpha_k f(\boldsymbol{x}_i|\theta_k)}.$$

Where g is given as

Incomplete supervision

- Human intervention

  - Active learning : To minimize the number of queries such that labeling cost can be minimized and model can be well generalized



Informativeness : how well an unlabeled instance helps reduce the un- certainty of a statistical model
ex) query-by-committee

Representativeness : how well an instance helps repre- sent the structure ofinput patterns
ex) clustering

Incomplete supervision

- Without Human intervention

  Semi-supervised learning : Cluster, manifold assumption

  - ✓ Generative methods : assume the labeled and unlabeled data are from same distribution, estimated by EM method

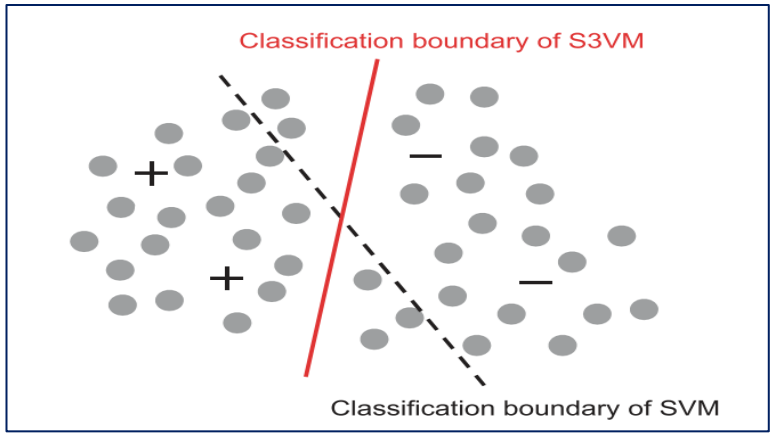
  - ✓ Graph-based methods : construct a graph, nodes correspond to instance, and edge to relation(similarity), then propagate label information

Incomplete supervision

- Without Human intervention

  Semi-supervised learning : Cluster, manifold assumption

  ✓ Low-density separation methods: classification boundary to go across the less-dense regions in input space



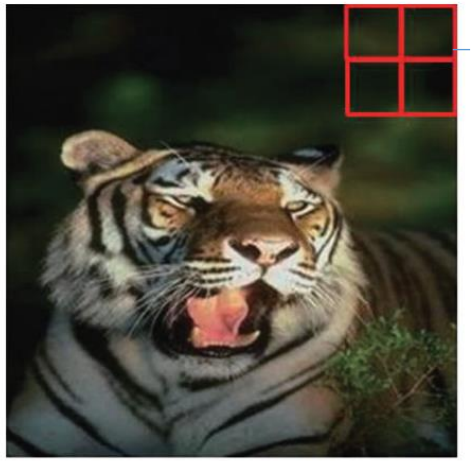
  ✓ Disagreement methods: Co-training

Inexact supervision

- Some supervision given, but not exactly as desired
  Multi-Instance Learning

motive

People have a key chain(bag) that contains few keys(instances). Some of these people are able to enter a certain room, and some aren't. The task is then to predict whether a certain key(instance) or a certain key chain(bag) can get you into that room. To solve this problem we need to find the exact key(instance) that is common for all the "positive" key chains(bag).



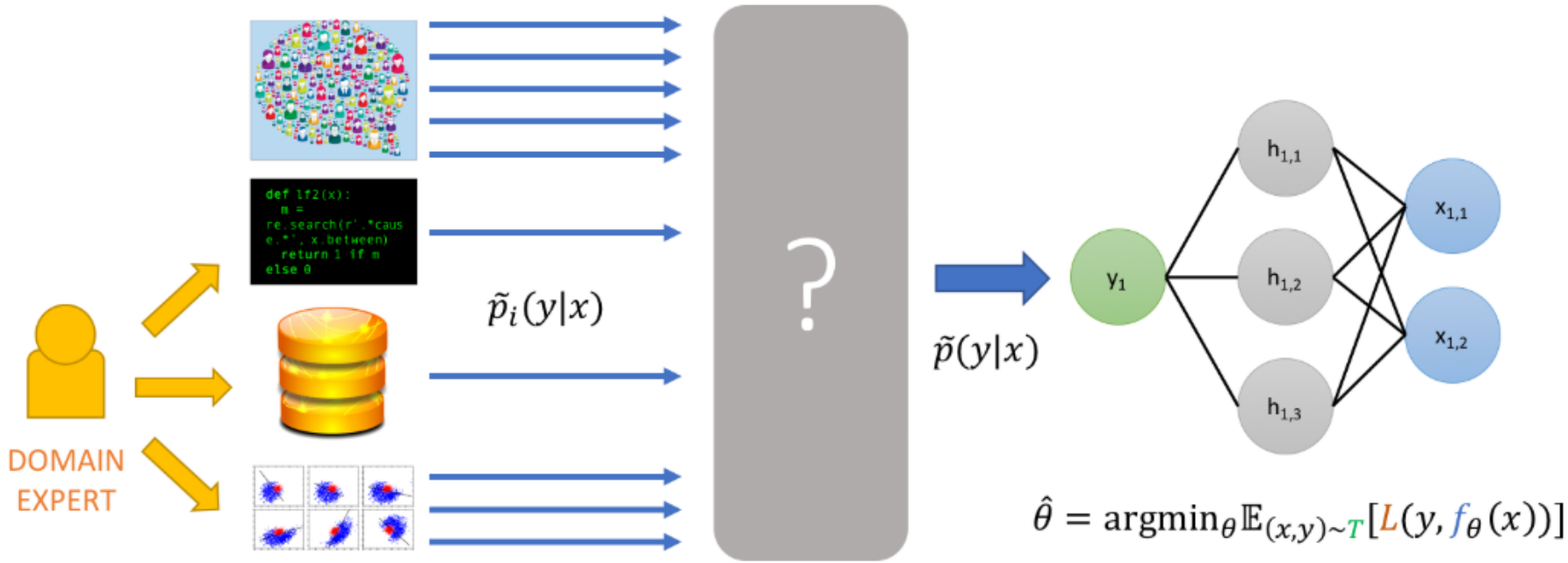
instances

bag

Predict unseen bag(the new image)

or

Identify the key instance(ROI)



Example Weak Supervision Sources

Technical Challenge: Integrating & Modeling Diverse Sources

Use Weak Supervision to Train End Model