Pessimistic Active Learning Using Robust Bias-Aware Prediction

Anqi Liu, Lev Reyzin, Brian Ziebart

UIC

Pool-Based Active Learning

- A pool based active learning algorithm [Lewis-Gale '94] sequentially chooses data-point labels to solicit from a pool of examples.
  - Usually constructs estimate of conditional label distribution $P(y|x)$ from labeled dataset.
  - Uses own estimate to select next datapoint label.

(this talk will focus on minimizing logloss, but ideas are more general)

Uncertainty Sampling

- Many active learning strategies employ uncertainty sampling – selecting examples about which the algorithm is least certain.
- Other strategies assess how a label:
  - is expected to change the prediction model [Settles-Craven ’08]
  - reduces an upper bound on the generalization error in expectation [Mackay ’92]
  - represents the input patterns of remaining unlabeled data [Settles ’12]

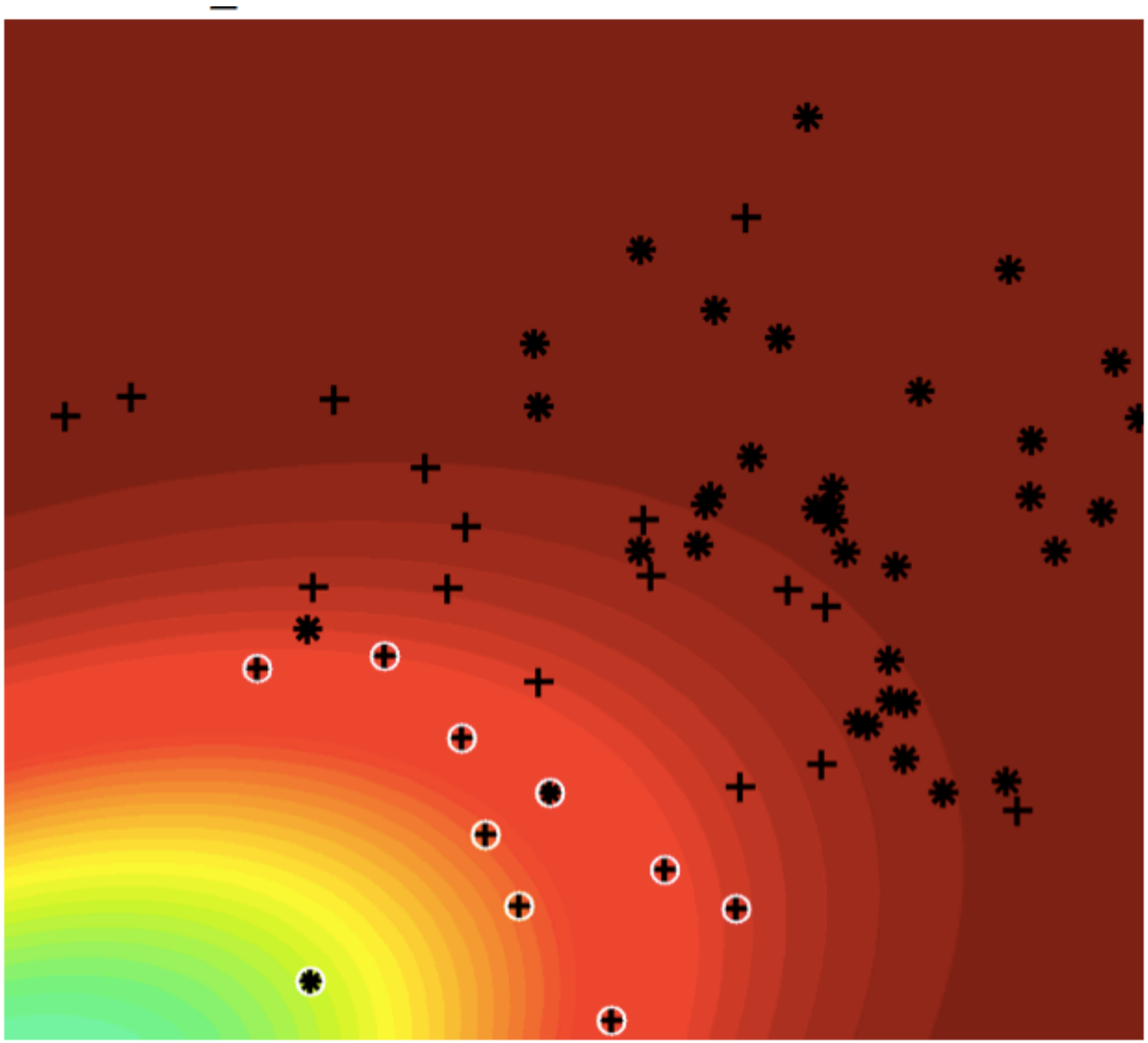
A Problem

Current active learning algorithms often perform poorly in practice [Attenberg-Provost ’11].

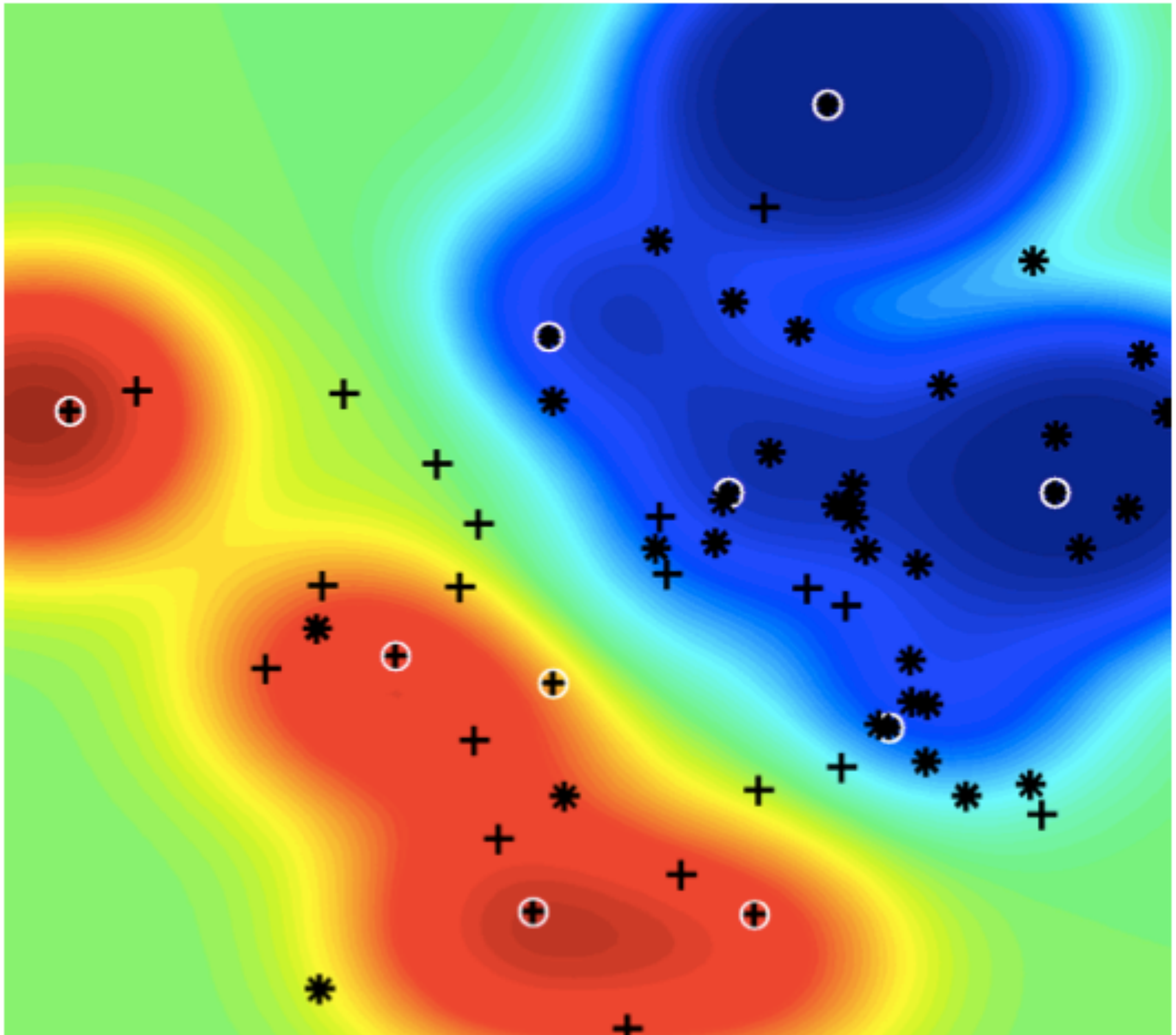
Why?

- In order to be take advantage of active learning, a biased label solicitation strategy should be used.
- Most current active learning strategies are overconfident, given this bias.

Typical Behavior of an Active Learner



Desired Behavior



Some Attempts to Fix This

- Seeding the active learner with a small random set [Dligach-Palmer ’11].
- Restricting the active learner to a small set of examples [Schein-Ungar ’07].
- Etc.

However, these modifications treat the symptoms of optimistic modeling and biased sampling and restrict the active learner, undermining its purported benefit.

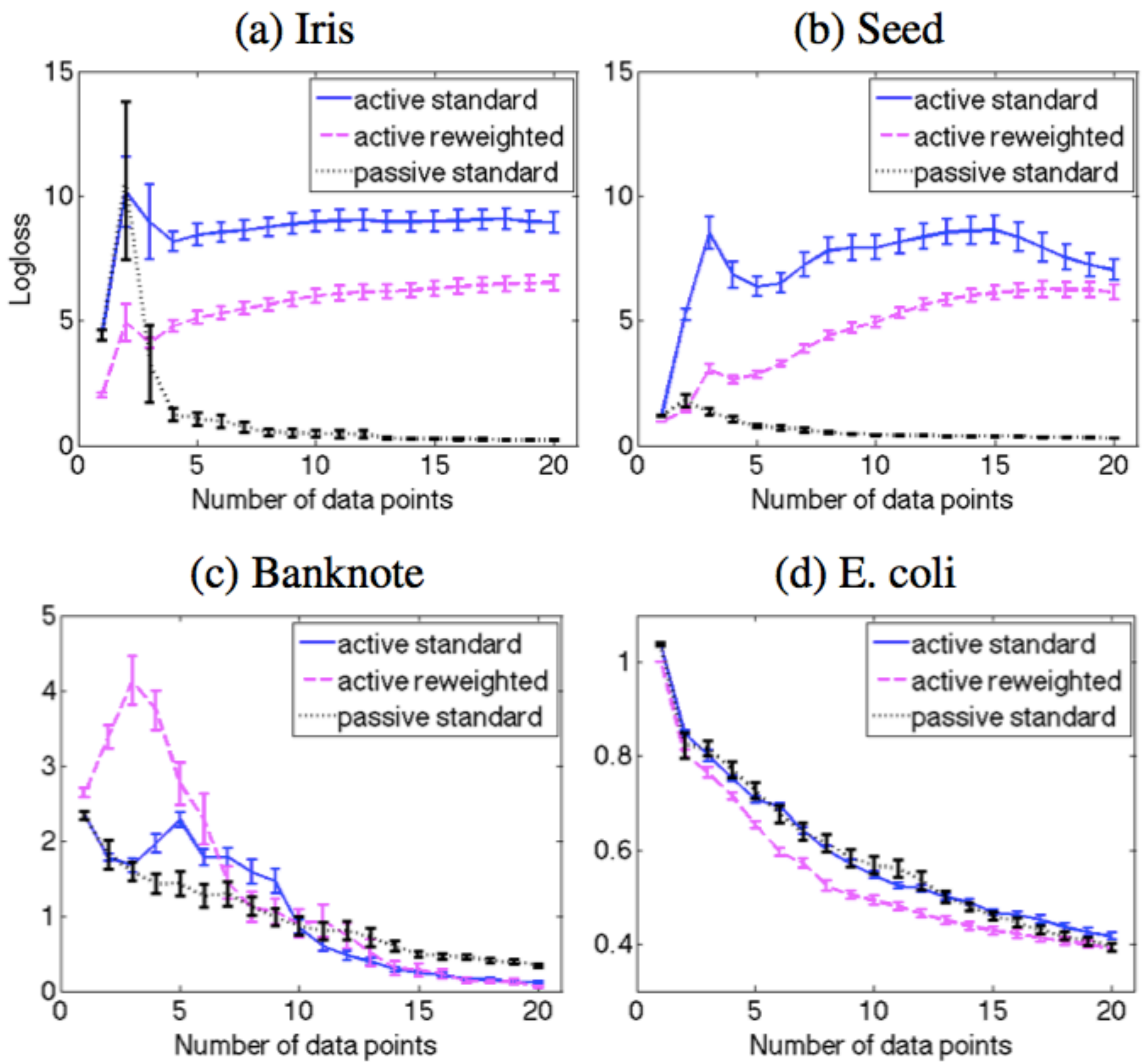
Biased Label Solicitation

When a non-uniform label-solicitation strategy is used, sample selection bias exists. In this case, it is known as covariate shift -- P(Y|X) is shared in source and target distributions.

Tackling covariate shift is difficult. For logistic regression, a common approach is importance re-weighting of source samples x according to $P_{trg}(x)/P_{src}(x)$ and minimizing a reweighted version of the target loss [Shimodaria '00].

Unfortunately this converges slowly [Cortes-Mansour-Mohri '10] and the variance of estimates is too high to be useful.

Logistic Regression Models



Approach

- We use the recently developed RBA (robust bias-aware prediction) framework for tackling covariate shift [Liu-Ziebart '14].
- RBA solves a game against a constrained adversary that chooses an evaluation distribution:

$$\min_{\hat{P}(y|x)} \max_{\check{P}(y|x) \in \tilde{\Xi}} \mathbb{E}_{P_{\mathcal{D}}(x)\check{P}(y|x)} \overbrace{[-\log \hat{P}(Y|X)]}^{\text{logarithmic loss}}$$

The set $\tilde{\Xi}$ constrains the adversary

Robust Prediction Strategy

- The RBA predictor can be obtained by solving the dual of a conditional max entropy estimation problem [Liu-Ziebart '14].

- Can be shown to upper bound the the generalization loss, under some assumptions. [Grunwald-Dawid '04]

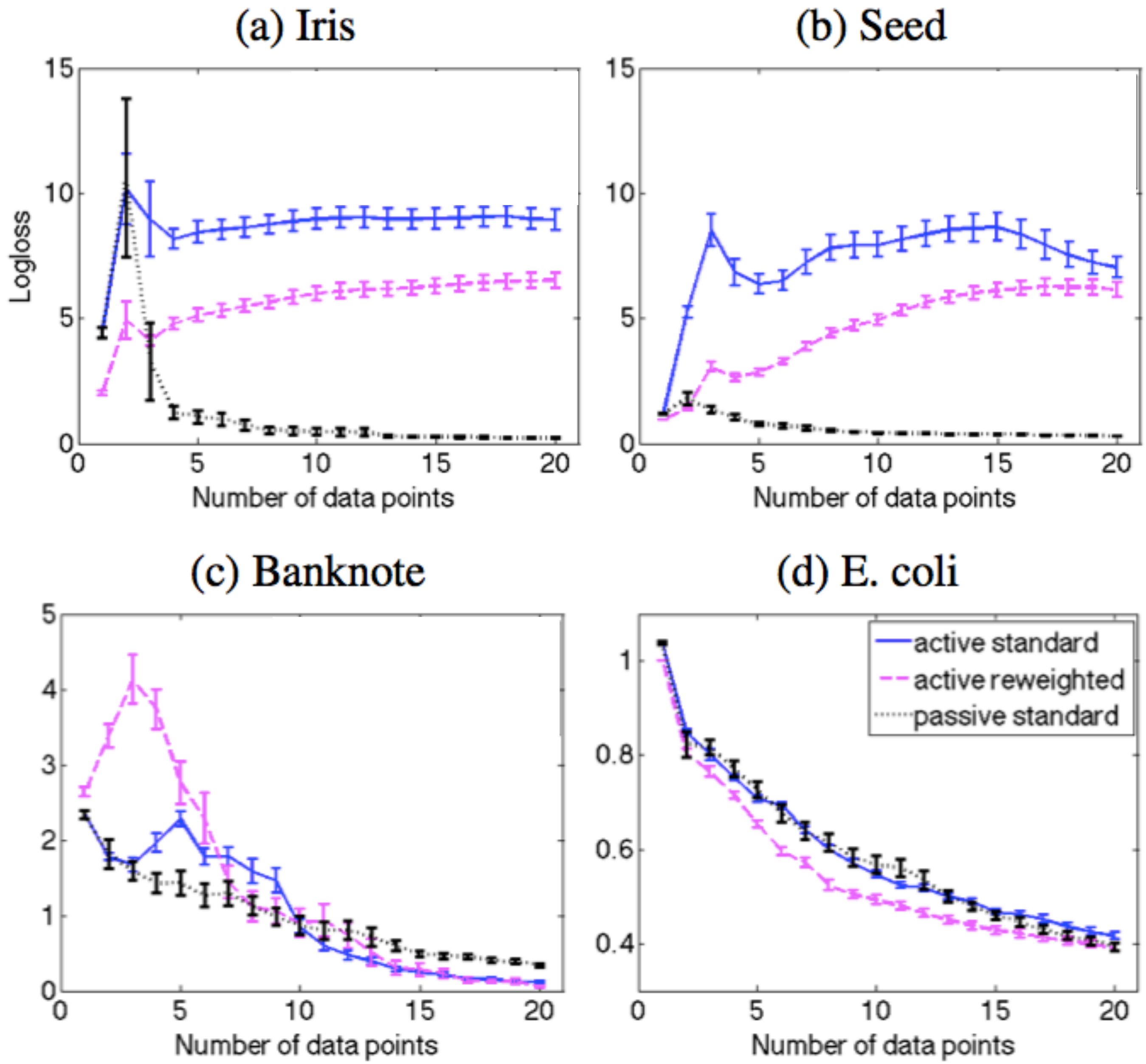
- $P_{src}(x)$ needs to be estimated – we use kernel density estimation with Gaussian kernels for $P_{src}(x)$.

- The RBA predictor turns out to less certain where the labeled data underrepresents the full data distribution.

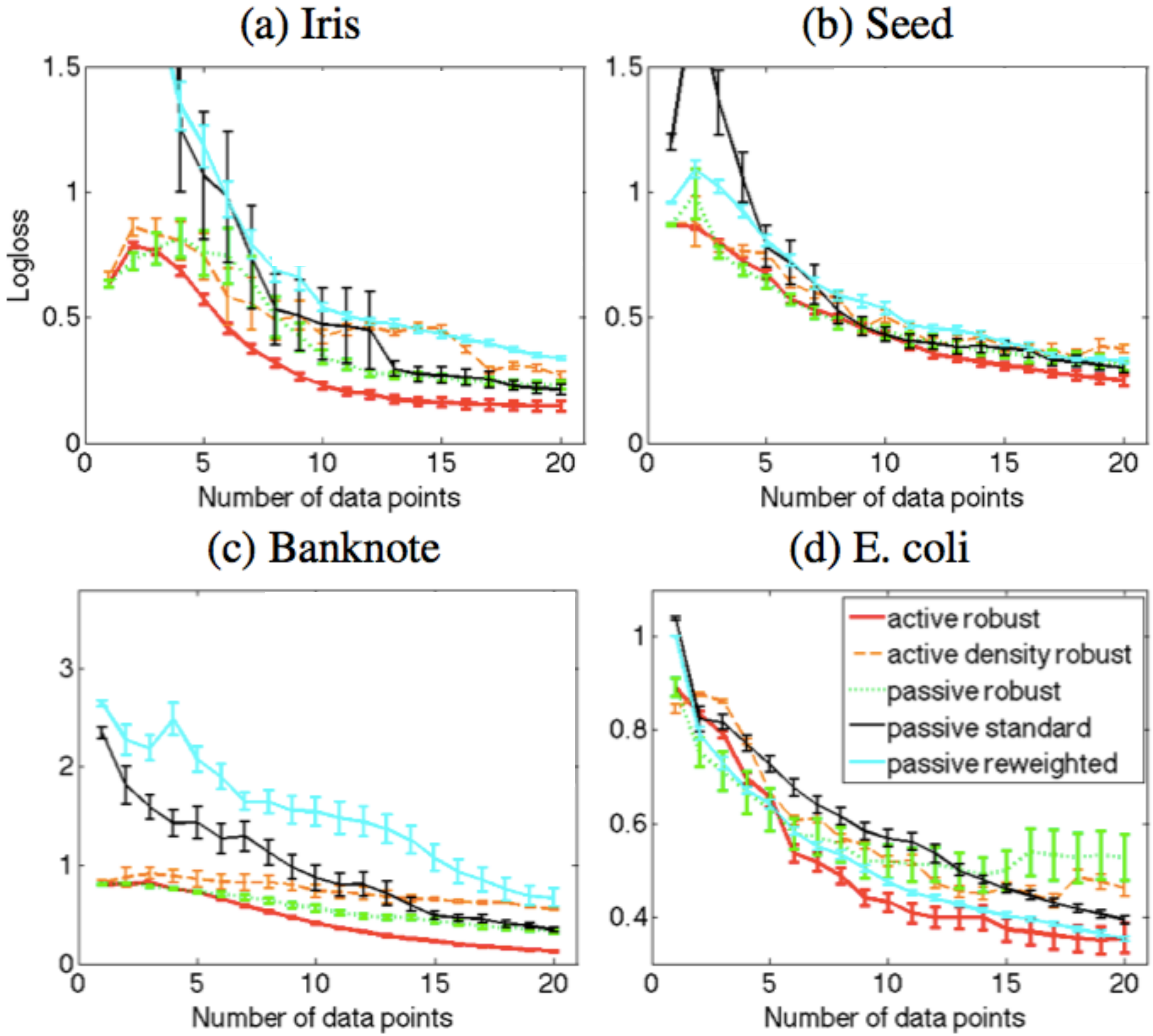
Sampling Strategies

- active robust – select point with largest value conditioned entropy
- active random – select point at random
- active density – select point with highest density ratio of $P_D(x)/P_L(x)$

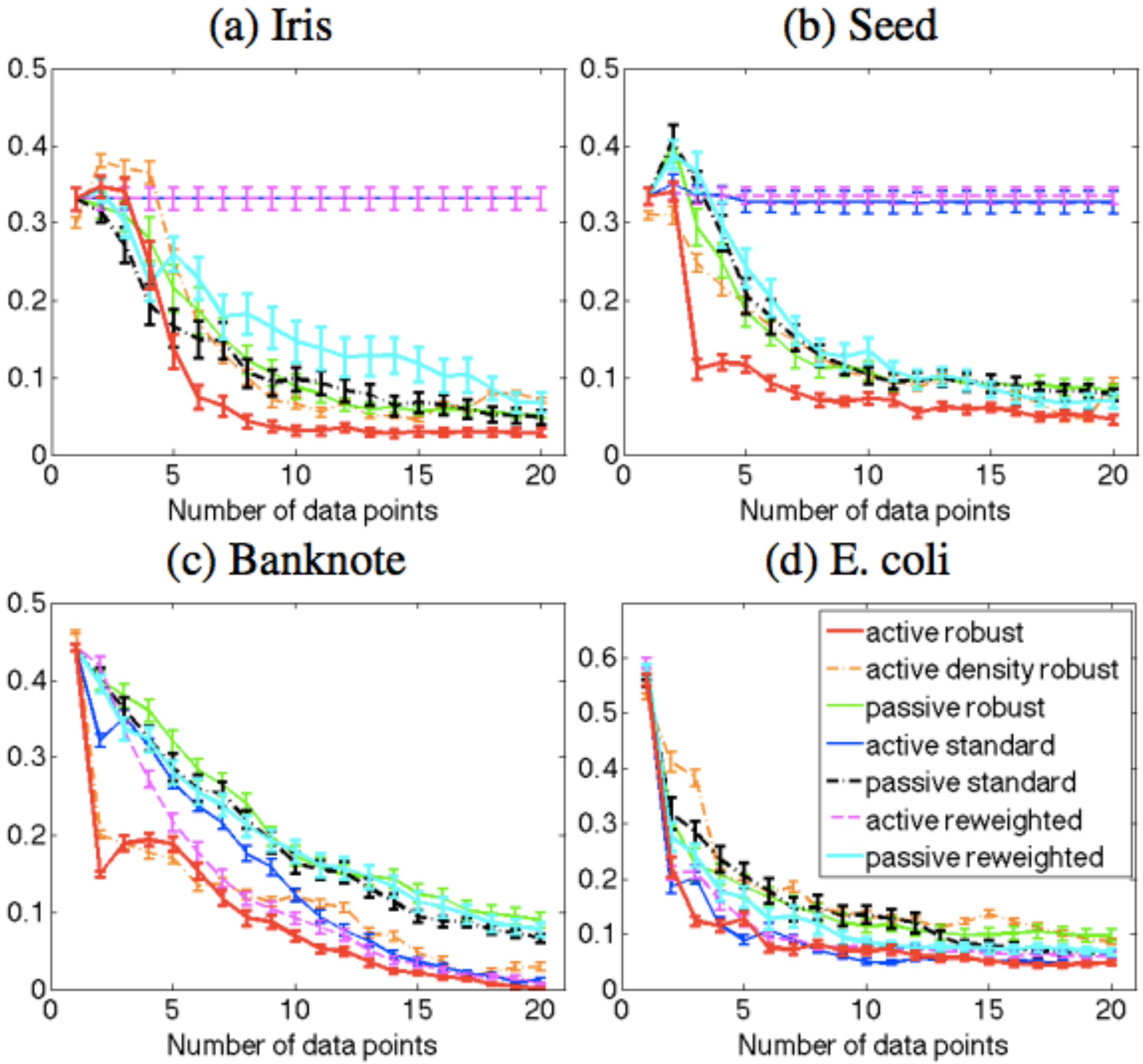
Standard Logistic Regression Models



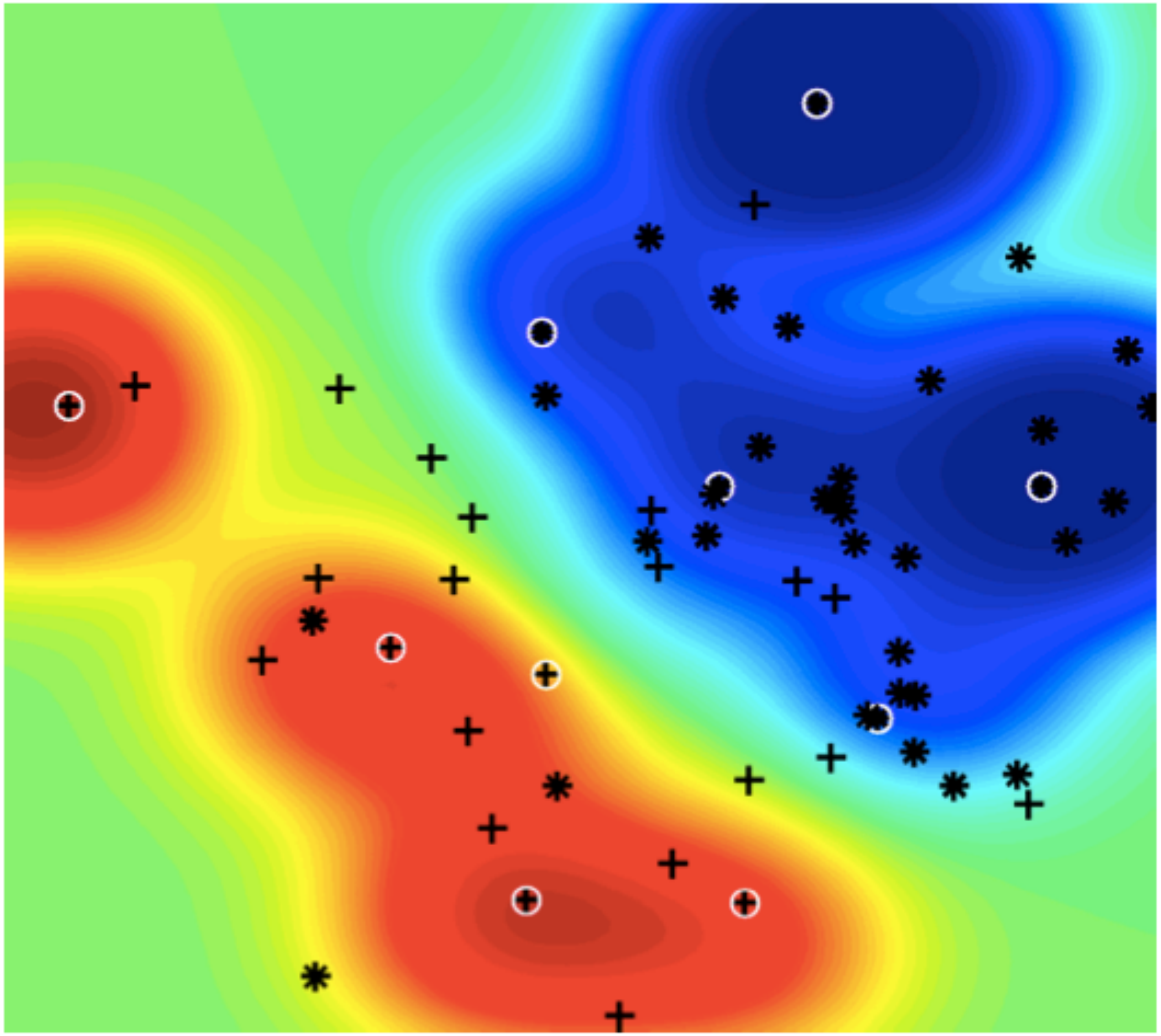
Our Results (logloss)



Our Results (classification error)



Predictions



Discussion

- Active learning inherently introduces covariate shift.
- Many active learners do not compensate for this properly or use unprincipled strategies.
- Recently developed techniques allow us to do robust active learning for logloss and beat many existing methods.
  - Even here, room for improvement.
- Other loss functions also can be tackled directly.
- More learning problems can be viewed from this framework.