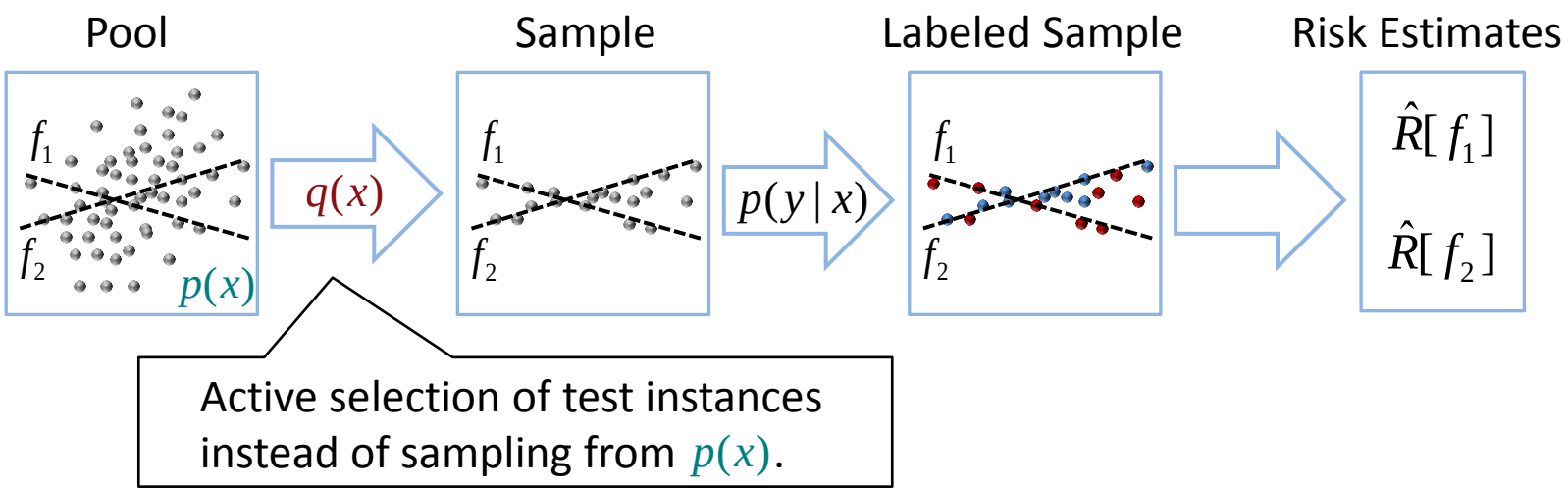
Active Comparison of Prediction Models

Poster: Tu25

Christoph Sawade, Niels Landwehr, and Tobias Scheffer

- **Given**:
  - Two candidate models $f_1(x), f_2(x)$
  - Pool of unlabeled test instances, that can be labeled at a cost.
- **Goal**: Identify model with lower risk $R[f] = \iint \ell(f(x), y)\, p(y \mid x)\, p(x) \mathrm{d}y\mathrm{d}x.$



Problem Setting

- Consistent risk estimate with instances drawn from $q(x) \neq p(x)$:

$$\hat{R}_{n,q}[f] = \left(\sum_{i=1}^{n} \frac{p(x_i)}{q(x_i)} \right)^{-1} \sum_{i=1}^{n} \frac{p(x_i)}{q(x_i)} \ell(f(x_i), y_i).$$

- Observed difference $\hat{\Delta}_{n,q} = \hat{R}_{n,q}[f_1] - \hat{R}_{n,q}[f_2]$ only random effect?

$$\Rightarrow \text{Apply statistical test with power } p(|\hat{\Delta}_{n,q}| > z_\alpha).$$

- **Goal**: Find instrumental distribution q^* that maximizes test power:

$$q^* = \arg\max_q \; p(|\hat{\Delta}_{n,q}| > z_\alpha).$$

Optimal Sampling Distribution

- **Lemma**: For sufficiently large n, test power is a monotonically decreasing function of estimator's variance $\mathrm{var}\left[\hat{\Delta}_{n,q}\right]$.

- Asymptotic variance:

$$n\,\mathrm{var}\left[\hat{\Delta}_{n,q}\right] \to \iint \frac{p(x)}{q(x)} \left(\ell(f_1(x), y) - \ell(f_2(x), y) - \Delta\right)^2 p(y \mid x)\, p(x)\, \mathrm{d}y\mathrm{d}x.$$

- **Theorem**: Instrumental distribution that asymptotically maximizes test power is given by:

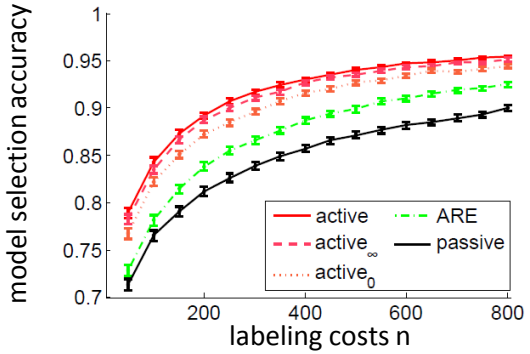
$$q^*(x) \propto p(x) \sqrt{\int \left(\ell(f_1(x), y) - \ell(f_2(x), y) - \Delta\right)^2 \mathrm{d}p(y \mid x)}.$$

Uniform distribution over pool.

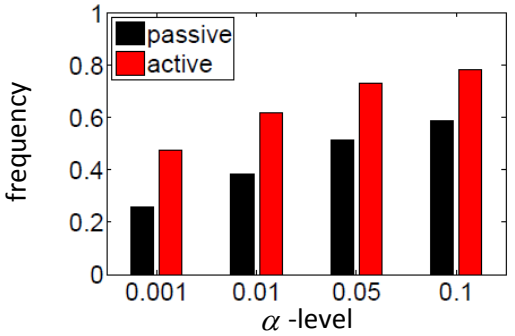
Can be approximated by the models.

Empirical Results

- Active comparison identifies model with lower risk more quickly than uniform sampling (costs reduced 60% - 90%).



- Active comparison made decisions more confidently (lower p-values).



- Active comparison does not lead to increased false-positive significance results.