Summary

Class prior shift heavily degrades the performance of positive-unlabeled classification (PU classification).

We propose two frameworks for solving this problem:
- Risk minimization framework
- Density ratio framework

We prove the equivalence of class prior shift and asymmetric error problems in PU classification.

Our methods are applicable for both problems!

Class prior shift

Positive-negative ratio in training and test data are different.

Decision boundary also shifts → Lead to low accuracy!

Practical examples: learn a classifier for a specific user from the internet and many users’ information.

PU classification

Given: Two sets of data

Positive: \[ X_+ := \{ x_i \} \sim \text{pos}(x) \]

Unlabeled: \[ X_U := \{ x_i \} \sim \text{pos}(x) + (1 - \pi_{\text{neg}}) \text{neg}(x) \]

Goal: Minimize either

Class prior shift classification risk:

\[
R^{\text{Shift}}_{\text{Class}}(g) = \pi_{\text{tr}} \text{Ep}[\pi_{\text{u}}(g(x)) + (1 - \pi_{\text{tr}}) \text{EN}[\pi_{\text{u}}(g(x))] \]

Asymmetric error classification risk:

\[
R^{\text{Aym}}_{\text{Class}}(g) = (1 - \alpha) \pi_{\text{tr}} \text{Ep}[\pi(g(x))] + \alpha (1 - \pi_{\text{tr}}) \text{EN}[\pi(g(x))] \]

\(\alpha \in (0, 1)\): false negative error

Existing PU classification work:

- no class prior shift, no asymmetric error (du Plessis+, 2015, Kiryo, 2017).

Equivalence of class prior shift and asymmetric error in PU classification

\[
\frac{\pi_{\text{tr}} \pi_{\text{u}}}{(1 - \pi_{\text{tr}}) \pi_{\text{u}}} = \frac{\pi_{\text{tr}} - \pi_{\text{tr}} \pi_{\text{u}}}{(1 - \pi_{\text{tr}}) \pi_{\text{u}} - 2 \pi_{\text{tr}} \pi_{\text{u}}}
\]

We can relate these problems based on the analysis of Bayes-optimal classifier.

Risk minimization approach

Risk:

\[
R^{\text{Shift}}_{\text{Class}}(g) = \pi_{\text{tr}} \text{Ep}[\pi_{\text{u}}(g(x)) + (1 - \pi_{\text{tr}}) \text{EN}[\pi_{\text{u}}(g(x))] \]

Using the following identity: \[ E_{\pi}[\cdot] = \pi_{\text{tr}} \text{Ep}[\cdot] + (1 - \pi_{\text{tr}}) \text{EN}[\cdot] \]

We can equivalently express the risk as

\[
R^{\text{Shift}}_{\text{Class}}(g) = E_{\pi}[\pi_{\text{u}}(g(x)) - \pi_{\text{tr}} \pi_{\text{u}} (g(x)) + \pi_{\text{tr}} \pi_{\text{u}} (g(x)) - \frac{\pi_{\text{tr}} \pi_{\text{u}}}{(1 - \pi_{\text{tr}}) \pi_{\text{u}}} - \text{EN}[\pi_{\text{u}}(g(x))] + \frac{1 - \pi_{\text{tr}} \pi_{\text{u}}}{(1 - \pi_{\text{tr}}) \pi_{\text{u}}} \text{EN}(\pi_{\text{u}}(g(x)))]
\]

Coincides with the existing method (du Plessis+, 2015) when \( R^{\text{Shift}}_{\text{Class}}(g) = R_{\text{tr}} \text{Ep}[\pi_{\text{u}}(g(x))] \)

Minimize empirical risk with surrogate loss (Bartlett+, 2006).

Density ratio approach

Bayes-optimal classifier:

\[
\text{unl}(x) = \pi_{\text{tr}} \text{pos}(x) + (1 - \pi_{\text{tr}}) \text{neg}(x)
\]

can be equivalently expressed as

\[
\text{unl}(x) = \pi_{\text{tr}} \text{pos}(x) + (1 - \pi_{\text{tr}}) \text{neg}(x)
\]

Another formulation: Density ratio!

\[
\text{unl}(x) = \frac{\pi_{\text{tr}} \pi_{\text{u}}}{(1 - \pi_{\text{tr}}) \pi_{\text{u}}} \text{pos}(x)
\]

Q: Which formulation is preferable?

In general, density ratio is bounded.

In PU classification, density ratio \( \frac{\pi_{\text{tr}} \pi_{\text{u}}}{(1 - \pi_{\text{tr}}) \pi_{\text{u}}} \) is bounded.

Lower and upper bounded

Naïve approach: estimate \( \text{unl}(x) \) and \( \text{unl}(x) \) separately then calculate \( \text{unl}(x) \) and \( \text{unl}(x) \)

Division operation amplifies the estimation error!

More effective direct approach:

unconstrained Least-squares Important Fitting (uLSIF) (Kanamori+, 2012).

Experiments

Datasets: banana, ijcnn1, MINIST, susy, cod-ma, magic

Methods:

- Density ratio \( \hat{\text{uLSIF}}, \hat{\text{uLSIF}} \)
- Linear-in input model: Double hinge (OH-Lin) and squared (Sq-Lin) losses. Kernel model (ker): Double hinge (OH-Ker) and squared (Sq-Ker) losses.

Parameter selection: (regularization, kernel width) 5-fold cross-validation.

Q: Which formulation is preferable?

In general, density ratio is bounded.

In PU classification, density ratio \( \frac{\pi_{\text{tr}} \pi_{\text{u}}}{(1 - \pi_{\text{tr}}) \pi_{\text{u}}} \) is bounded.

In general, density ratio is bounded.

In PU classification, density ratio \( \frac{\pi_{\text{tr}} \pi_{\text{u}}}{(1 - \pi_{\text{tr}}) \pi_{\text{u}}} \) is bounded.

Lower and upper bounded

Naïve approach: estimate \( \text{unl}(x) \) and \( \text{unl}(x) \) separately then calculate \( \text{unl}(x) \) and \( \text{unl}(x) \)

Division operation amplifies the estimation error!

More effective direct approach:

unconstrained Least-squares Important Fitting (uLSIF) (Kanamori+, 2012).

Experiments

Datasets: banana, ijcnn1, MINIST, susy, cod-ma, magic

Methods:

- Density ratio \( \hat{\text{uLSIF}}, \hat{\text{uLSIF}} \)
- Linear-in input model: Double hinge (OH-Lin) and squared (Sq-Lin) losses. Kernel model (ker): Double hinge (OH-Ker) and squared (Sq-Ker) losses.

Parameter selection: (regularization, kernel width) 5-fold cross-validation.

Results reported in mean and std. of accuracy of 10 trials.

Dataset information and more experiments and can be found in the paper.

References