

Match quality when prognosis and propensity are highly correlated

Rachael Caelie (Rocky) Aikens

4/23/2019

Set Up

We compare the performance of propensity score matching, Mahalanobis distance matching, and Buffalo Matching (described in the previous section) on simulated data, varying the dimensionality of the problem, the fixed treatment to control ratio during matching, and the correlation between the true propensity and prognostic score. The generative model for all of our simulations is the following:

$$\begin{aligned}X_i &\sim_{iid} \text{Normal}(0, I_p), \\T_i &\sim_{iid} \text{Bernoulli}\left(\frac{1}{1 + \exp(-\phi(X_i))}\right), \\Y_i &= \tau T_i + \Psi(X_i) + \epsilon_i, \\ \epsilon_i &\sim_{iid} N(0, \sigma^2),\end{aligned}$$

where the true propensity and prognostic scores are given by the linear combinations

$$\begin{aligned}\phi(X_i) &= X_{i1} - 10/3, \\ \Psi(X_i) &= \rho X_{i1} + \sqrt{(1 - \rho^2)} X_{i2},\end{aligned}$$

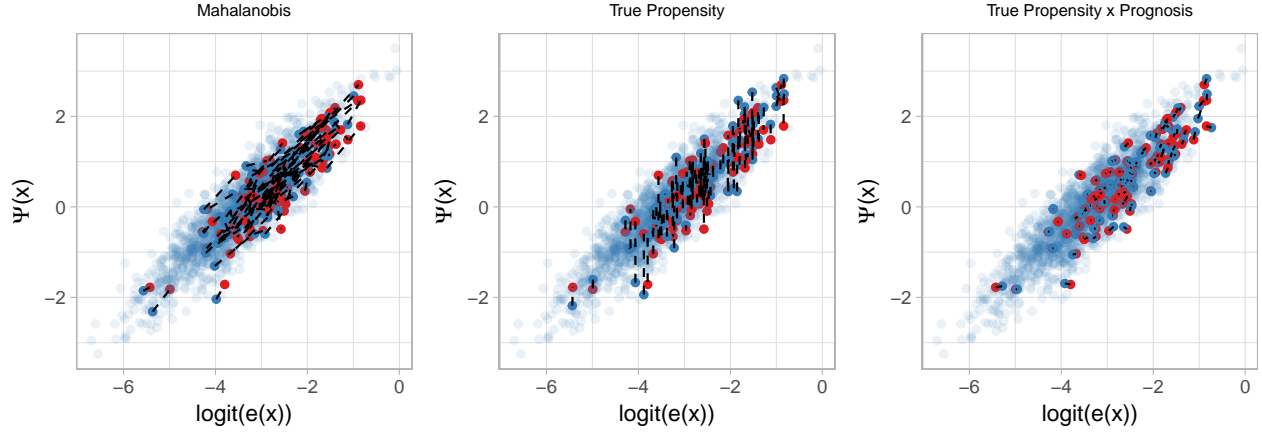
so that $\text{Cor}(\phi(X_i), \Psi(X_i)) \propto \rho$. The propensity score formula was chosen such that there were approximately 100 treated observations in each dataset. We consider $p = 10$, $\rho = 0, 0.1, \dots, 0.9, 1.0$, and $k = 1, \dots, 10$. Each simulation consisted of a dataset of size $n = 2000$ and was repeated $N = 1000$ times. We fix the treatment effect to be constant with $\tau = 1$ and the noise to be $\sigma = 1$. For a given matching, we estimate ATT and design sensitivity $\tilde{\Gamma}$ using the permutation t -statistic from the package `sensitivitymv`

Replication of Dylan's Plots

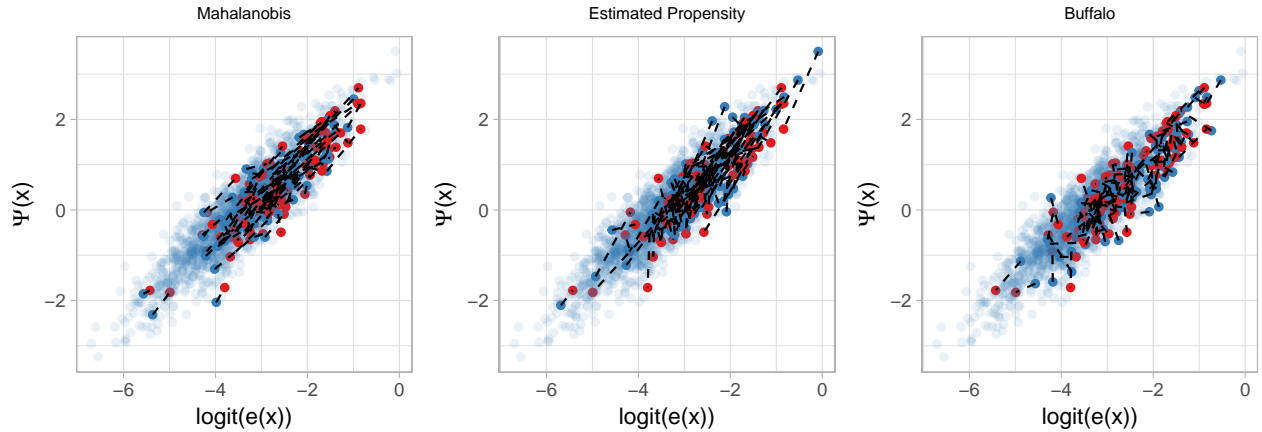
Basic Visualization

Here, I assume propensity and prognostic information are highly correlated $\rho = 0.9$.

Below, we imagine that we know the true propensity and prognostic scores, and we match on those for buffalo and propensity score matching. Unsurprisingly, life is very good for buffalo matchers when the true scores are known.



Now, we take the same data set and imagine that we don't know the true propensity and prognostic scores, and so must build them empirically. For propensity score matching, we build a logistic regression of treatment assignment on all of the variables, using the entire dataset. For buffalo, we use the propensity score logistic regression in addition to a prognostic score, which we fit on a subset of the controls that are chosen to be good matches to the treated individuals based on mahalanobis distance.



We made the observation that variance actually increased with k when ρ was large. One explanation that we thought of was that, when ρ is large, propensity and prognosis are highly correlated; As we match more and more control individuals to each individual *all* methods are selecting matches which are more distant in prognosis and in propensity. The propensity distance increases the bias of the estimator, while the prognosis distance increases the variance of the estimator. In the visualization plots above, we only show 1:1 matching.

I would predict that this issue is a bigger deal when X_1 is more strongly predictive of treatment assignment (i.e. when treatment assignment is less random and more based on covariates), since this problem is mostly an issue in the regions of limited overlap. This would explain why we saw variance increase with k when $\text{logit}(e(x)) = X_1 - 10/3$, but not when $\text{logit}(e(x)) = X_1/3 - 4/3$.