Estimating Propensity Scores and Testing Overlap using Support Vector Machines

Ari Boyarsky

Advisor: Prof. Alex Torgovitsky (Department of Economics)

- Motivation
  - Matching methods are common across the social sciences. They are powerful but require strong assumptions, in particular that,

    $$\mathbb{E}[Y_0|D=1,X] = \mathbb{E}[Y_0|D=0,X] \tag{1}$$

    Heckman et al. (1998) and region of common support,

    $$S = \mathrm{Supp}(X|D=1) \cap \mathrm{Supp}(X|D=0) \tag{2}$$

  - Propensity score matching attempts to simplify this. Using $p(X) := P[D=1|X]$ (usually with logistic regression). In this case, overlap requires that,

    $$p(X) \in (0,1) \tag{3}$$

    Such that treatment outcome is not perfectly predicted.
    - Heinrich et al. (2010), Harder et al. (2011), Imbens (2014)
  - But, these methods may misidentify overlap (i.e., a nonlinear boundary). This paper provides a test to resolve this issue.

Proposed Methodology

- If we want to be sure that we have common support we should employ a methodology that maximizes predictive power.
- Using Support Vector Machines (SVM) with radial basis function kernels allows us to find nonlinear boundaries in a high dimensional space and assess the fit using standard machine learning model estimates. Primal problem,

$$\begin{aligned}
\min_{w} \quad & \frac{1}{2}\langle\, w, w\rangle + C\sum_{i=1}^{N}\xi_i && \text{(4a)}\\
\text{s.t.} \quad & y_i(\langle\, w, \phi(x_i)\rangle + b) \geq 1 - \xi_i && \text{(4b)}\\
& \xi_i \geq 0 \;\; \forall i = 1, \ldots, N && \text{(4c)}
\end{aligned}$$

$$K(u, v) = \langle\, \phi(u), \phi(v)\rangle = e^{-\gamma||u-v||_2^2} \tag{5}$$

- K-fold cross validation error to evaluate fit.

Proposed Methodology

- We can also use Platt (1999) transforms to calculate $p(X)$,

$$P(y = 1|f_i) = p(x_i) = \frac{1}{1 + exp(Af(x_i) + B)} \tag{6}$$

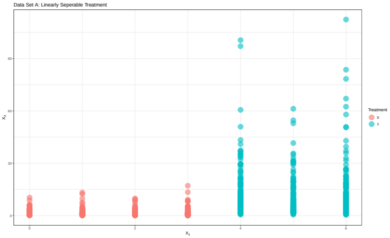
A and B are fit with maximum likelihood estimation using the initial data set,

$$\min_{A,B} - \sum_{i}^{N} \frac{y_i + 1}{2} \log(p(x_i)) + (1 - \frac{y_i + 1}{2}) \log(1 - p(x_i)) \tag{7}$$

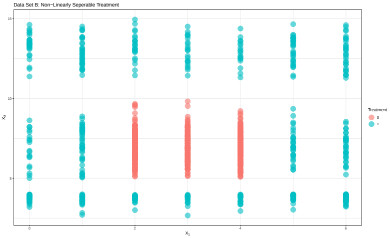
  - We should pay particular attention to observations close to the separating hyperplane i.e. $p(X) \in [0.4, 0.6]$.
    - Trade off between within sample accuracy and generality.
  - If we have an imbalanced data set we can use weights to adjust for this (like in LaLonde (1986)).

Simulations of Test Data

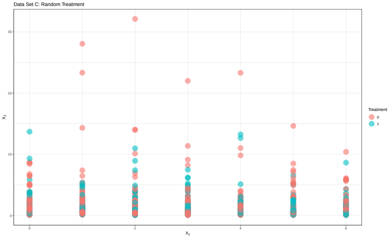
(a) Convex Seperation



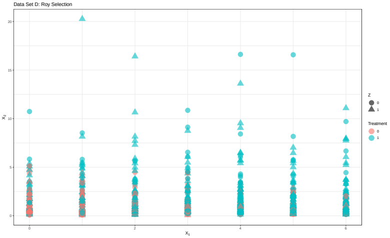
(b) Nonlinear Seperation



(c) Random Treatment



(d) Roy Selection



Simulational Results: SVM

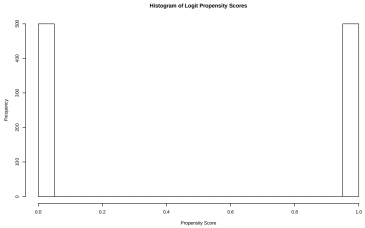
Table: SVM Overlap Tests of Test Data Sets (A)-(D)

Data Set	(A)	(B)	(C)	(D)
% Fixed Cross Validation Error	0.210	0.196	33.133	16.190
$K = 10$ CV Accuracy Rate	99.6	100	56.6	74.5
Brier Score	0.0	0.0	0.48	0.17
% Margin (0.1 Width)	0	0	0	5.5
% Margin (0.2 Width)	0	0	80.7	14.5
Kernel	Radial	Radial	Radial	Radial
Gamma	1	2	1	0.01
Cost	32	256	4	4
N	1,000	1,000	1,000	1,000

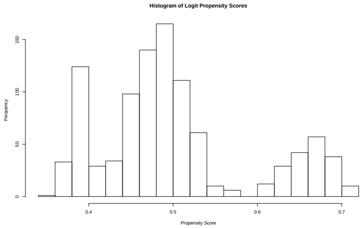
Note: Fixed Cross Validation Error is computed by cross validating the entire data set against gamma values $\{0.01, 0.1, 0.5, 1, 1.5, 2\}$ and cost parameters $\{4, 8, 16, 32, 64, 128, 256, 512, 1024\}$ using K-fold cross validation error with a fixed test and validation sample.

Simulational Results: Logistic Regression

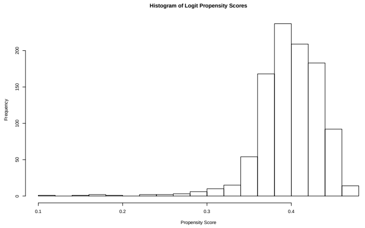
(a) Convex Seperation



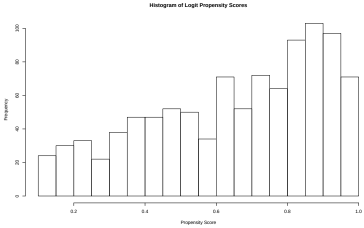
(b) Nonlinear Seperation



(c) Random Treatment

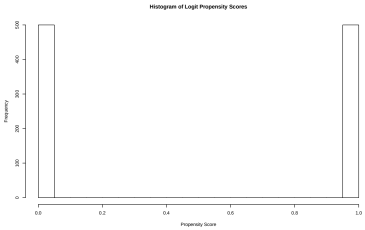


(d) Roy Selection

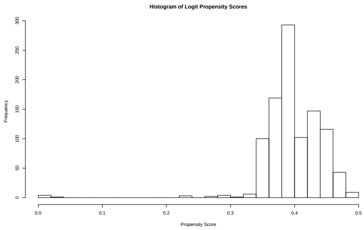


Simulational Results: Logistic Regression (Higher Order Terms)

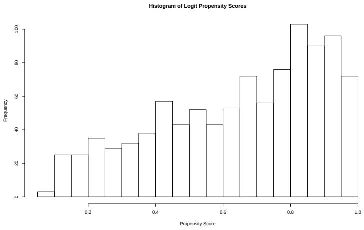
(b) Nonlinear Seperation



(c) Random Treatment



(d) Roy Selection



Application: LaLonde (1986) Data

- Dehejia and Wahba (1999) – Issues with sample and variable selection (Smith and Todd (2005))
- Could there also be issues with overlap?
  - DW (1999): Compute $p(X)$, assess overlap/trim data to increase overlap, match.
  - We follow the same procedure but use an SVM.
  - After trimming such that $p(X) \in (0.05, 1)$ we have only 178 observations, the same specification with logistic regression yields 561 observations.
  - The data set is unbalanced 118 treated, only 49 control, 146 observations with propensity scores in $0.7 - 0.72$ is suggestive of overlap.
  - However, we should weight data to address imbalance. So,

$$G(x) = sgn[wf(x) + b] \tag{8}$$

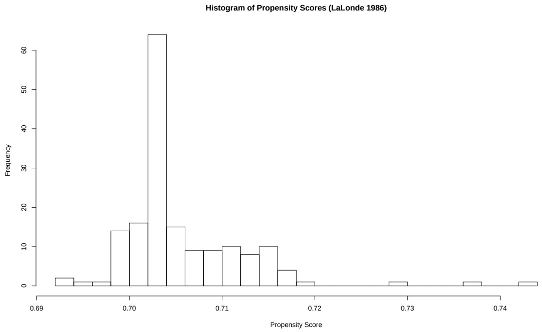
Where w is vector of inverse frequency weights ($k = 2$ classes),

$$w_j = \frac{n}{kn_j} \tag{9}$$

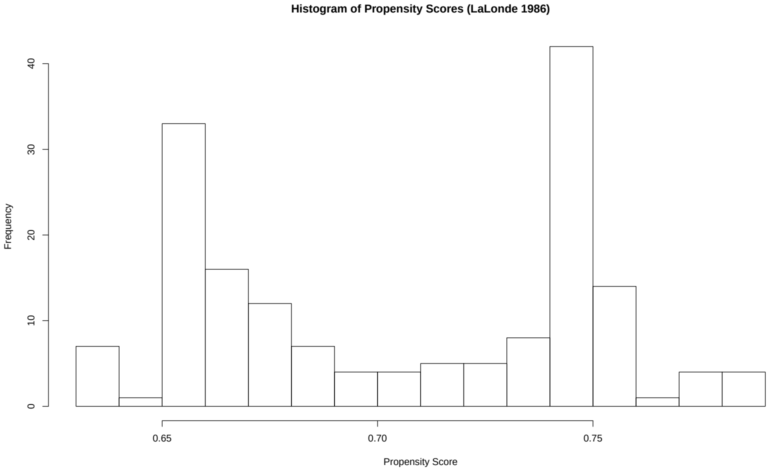
LaLonde (1986) SVM Results (before weighting)

	(1)
% Cross Validation Error	0.1984
% Margin (0.1 Width)	0
% Margin (0.2 Width)	0
Kernel	Radial
Gamma	0.5
Cost	4
N	178

Note: Training Cross Validation Error is computed by cross validating the entire data set against gamma values $\{0.01, 0.1, 0.5, 1\}$ and cost parameters $\{4, 8, 16, 32, 64, 128, 256\}$ using the K-fold cross validation such that we split the data into K subsets, train the SVM on every other subset, and compute the average test error for each subset. This specification predicts D using: age, education, no degree, black, married, Hispanic, RE74, RE75, u74, u75, education*RE74.



LaLonde (1986) SVM Results (after weighting)



Note: Applying inverse frequency weighting yields a control weight of 6.816, and a treated weight of 2.831.

So there is overlap, but in a much smaller support than considered by DW (1999). ATT of about 1643.908 matching Imbens (2014).