# Preprocessing with SVM

Ryan Del Bel

March 6, 2014

We first consider removing the contaminated observations by hand. Any reference to SVM or SV in this table referrs to removing the contaminated observatins by hand, and the contaminated observations themselves.

```
latex(tbl2, file = "", cgroup = c("", "AUC", "TOP10", "NULLP", ""), rgroup = c("1\\%",
    "2\\%", "5\\%", "10\\%"), n.cgroup = c(1, 2, 2, 2, 3), n.rgroup = c(5, 5),
    here = T, rowlabel = "", rowname = NULL)
```

| | **AUC** | | **TOP10** | | **NULLP** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| OR | SVM | Normal | SVM | Normal | SVM | Normal | #SV | #Con | %ConSV |
| 1.25 | 0.896 | 0.893 | 5.88 | 5.75 | 0.479 | 0.478 | 10 | 10 | 1 |
| 1.5 | 0.975 | 0.977 | 8.00 | 7.75 | 0.470 | 0.471 | 10 | 10 | 1 |
| 1.75 | 0.996 | 0.996 | 9.00 | 9.00 | 0.472 | 0.471 | 10 | 10 | 1 |
| 1.1-1.5 | 0.886 | 0.876 | 5.38 | 5.50 | 0.484 | 0.484 | 10 | 10 | 1 |
| 1.1-2 | 0.905 | 0.906 | 6.88 | 6.75 | 0.479 | 0.480 | 10 | 10 | 1 |
| 1.25 | 0.895 | 0.838 | 4.88 | 3.38 | 0.479 | 0.483 | 50 | 50 | 1 |
| 1.5 | 0.987 | 0.963 | 8.00 | 6.50 | 0.471 | 0.471 | 50 | 50 | 1 |
| 1.75 | 0.993 | 0.973 | 8.62 | 6.88 | 0.468 | 0.469 | 50 | 50 | 1 |
| 1.1-1.5 | 0.869 | 0.819 | 5.50 | 4.38 | 0.482 | 0.476 | 50 | 50 | 1 |
| 1.1-2 | 0.883 | 0.847 | 6.88 | 6.25 | 0.489 | 0.494 | 50 | 50 | 1 |

We see that when removing the contaminated points by hand we obtain better results.
The following table is a comprehensive look at the performance of SVM when contaminating observations and removing the 'worst' support vectors.

```
latex(tbl, file = "", cgroup = c("", "AUC", "TOP10", "NULLP", ""), rgroup = c("1\\%",
    "2\\%", "5\\%", "10\\%"), n.cgroup = c(1, 2, 2, 2, 3), n.rgroup = c(5, 5,
    5, 5, 5, 5), here = T, rowlabel = "", rowname = NULL)
```

| | AUC | | TOP10 | | NULLP | | | | |
|---|---|---|---|---|---|---|---|---|---|
| OR | SVM | Normal | SVM | Normal | SVM | Normal | #SV | #Con | %ConSV |
| 1.25 | 0.876 | 0.895 | 4.50 | 5.00 | 0.424 | 0.489 | 96.38 | 10 | 0.250 |
| 1.5 | 0.971 | 0.972 | 7.12 | 7.38 | 0.420 | 0.457 | 49.50 | 10 | 0.162 |
| 1.75 | 0.994 | 0.996 | 8.50 | 8.62 | 0.440 | 0.463 | 28.12 | 10 | 0.150 |
| 1.1-1.5 | 0.793 | 0.806 | 4.25 | 4.50 | 0.426 | 0.486 | 81.62 | 10 | 0.288 |
| 1.1-2 | 0.888 | 0.898 | 7.12 | 7.00 | 0.465 | 0.488 | 34.25 | 10 | 0.075 |
| 1.25 | 0.806 | 0.818 | 2.75 | 3.62 | 0.413 | 0.484 | 103.12 | 50 | 0.225 |
| 1.5 | 0.951 | 0.959 | 6.50 | 6.50 | 0.421 | 0.477 | 77.38 | 50 | 0.242 |
| 1.75 | 0.985 | 0.986 | 8.12 | 8.12 | 0.436 | 0.477 | 53.88 | 50 | 0.248 |
| 1.1-1.5 | 0.789 | 0.799 | 3.50 | 3.50 | 0.414 | 0.493 | 106.88 | 50 | 0.268 |
| 1.1-2 | 0.855 | 0.866 | 5.62 | 5.25 | 0.435 | 0.486 | 71.62 | 50 | 0.248 |
| 1.25 | 0.874 | 0.875 | 4.38 | 4.50 | 0.478 | 0.487 | 10.00 | 10 | 0.000 |
| 1.5 | 0.989 | 0.989 | 8.38 | 8.50 | 0.452 | 0.463 | 10.00 | 10 | 0.062 |
| 1.75 | 0.997 | 0.997 | 9.25 | 9.12 | 0.471 | 0.479 | 10.00 | 10 | 0.038 |
| 1.1-1.5 | 0.867 | 0.866 | 5.50 | 5.38 | 0.464 | 0.477 | 10.00 | 10 | 0.038 |
| 1.1-2 | 0.919 | 0.919 | 6.38 | 6.38 | 0.472 | 0.479 | 10.00 | 10 | 0.075 |
| 1.25 | 0.815 | 0.818 | 3.62 | 3.62 | 0.460 | 0.471 | 10.00 | 50 | 0.030 |
| 1.5 | 0.943 | 0.942 | 7.38 | 7.38 | 0.462 | 0.474 | 10.00 | 50 | 0.040 |
| 1.75 | 0.983 | 0.984 | 8.38 | 8.38 | 0.459 | 0.472 | 10.00 | 50 | 0.048 |
| 1.1-1.5 | 0.838 | 0.842 | 4.12 | 4.00 | 0.470 | 0.479 | 10.00 | 50 | 0.028 |
| 1.1-2 | 0.913 | 0.911 | 5.50 | 5.50 | 0.476 | 0.485 | 10.00 | 50 | 0.048 |
| 1.25 | 0.810 | 0.809 | 2.88 | 2.88 | 0.478 | 0.487 | 10.00 | 50 | 0.018 |
| 1.5 | 0.967 | 0.970 | 6.88 | 7.00 | 0.475 | 0.483 | 10.00 | 50 | 0.045 |
| 1.75 | 0.959 | 0.960 | 7.75 | 7.88 | 0.471 | 0.480 | 10.00 | 50 | 0.050 |
| 1.1-1.5 | 0.878 | 0.879 | 5.12 | 4.88 | 0.455 | 0.489 | 48.50 | 10 | 0.200 |
| 1.1-2 | 0.909 | 0.915 | 6.75 | 7.00 | 0.450 | 0.482 | 40.75 | 10 | 0.175 |
| 1.25 | 0.856 | 0.856 | 3.75 | 4.25 | 0.443 | 0.476 | 50.00 | 50 | 0.145 |
| 1.5 | 0.962 | 0.970 | 7.12 | 7.25 | 0.437 | 0.475 | 50.00 | 50 | 0.148 |
| 1.75 | 0.973 | 0.971 | 7.62 | 7.75 | 0.428 | 0.465 | 46.88 | 50 | 0.215 |
| 1.1-1.5 | 0.795 | 0.803 | 4.88 | 4.62 | 0.443 | 0.480 | 50.00 | 50 | 0.148 |
| 1.1-2 | 0.846 | 0.855 | 5.38 | 5.75 | 0.451 | 0.482 | 49.12 | 50 | 0.165 |

We see that we do not obtain better results when removing the support vectors. Only a very small proportion of our removed support vectors are the comtaminated observations.