Calibrated probability estimation

Qimin Chen
A53284263
qic003@ucsd.edu

1 SVM training procedure

Figure 1 shows the cross-validated error estimate and test set error rate of different valus C. Table 1 shows the accurate numerical values of cross-validated error estimate and test set error rate of different valus C. It is apparent that SVM has the smallest cross-validated error and test set error when $C = 0.1$. The value of C I end up using is 100000 since both cross-validated error and test set error go up after $C = 0.1$.

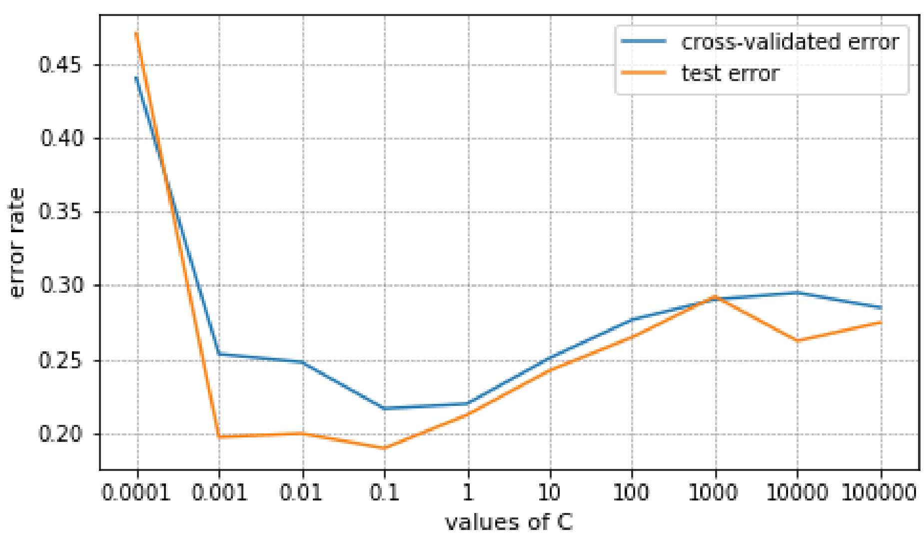

Figure 1: error rate versus different values of C

values of C	cross-validated error	test set error
0.0001	0.43999	0.47000
0.001	0.25368	0.19750
0.01	0.24820	0.20000
0.1	**0.21682**	**0.19000**
1	0.22002	0.21250
10	0.25092	0.24250
100	0.27683	0.26500
1000	0.29048	0.29250
10000	0.29502	0.26250
100000	0.28501	0.27500

Table 1: numerical value of cross-validated error and test set error of different values of C

2 Squashing function

Figure 2 shows the reliability diagram on the test set for Squashing function. It can be seen from figure that it is not calibrated because the predicted probability in each interval is way far from the perfectly calibrated probability which is the middle value of that interval.

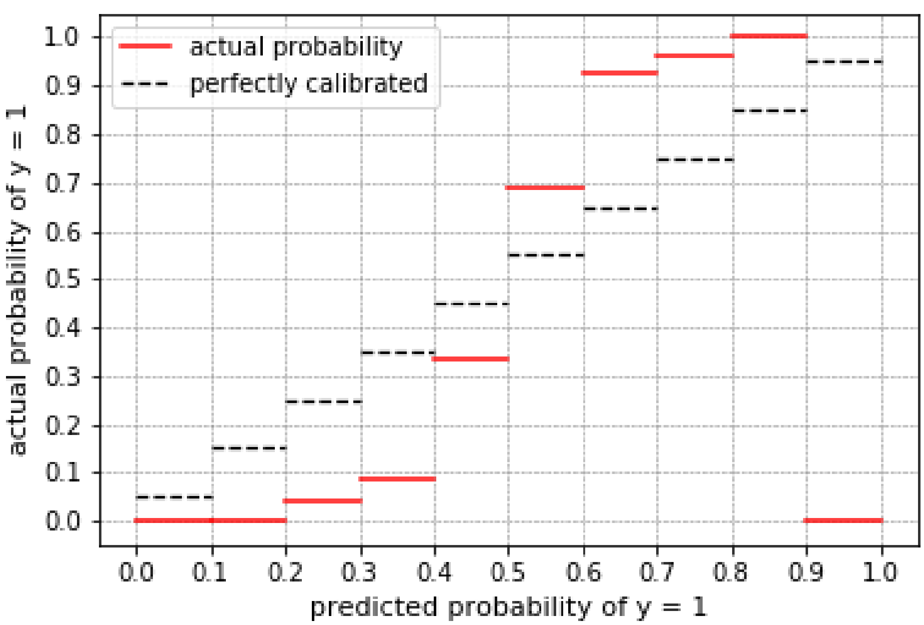

Figure 2: error rate versus different values of C

3 Platt scaling

Figure 3 shows the reliability diagrams of Platt scaling forcing $b = 0$ (left) and learning b (right). It can be seen from the figure that platt scaling with learning b is better than $b = 0$ because b is the learning bias that adjust the predicted probability getting closer to perfectly calibrated probability. This is better calibrated than option 1.

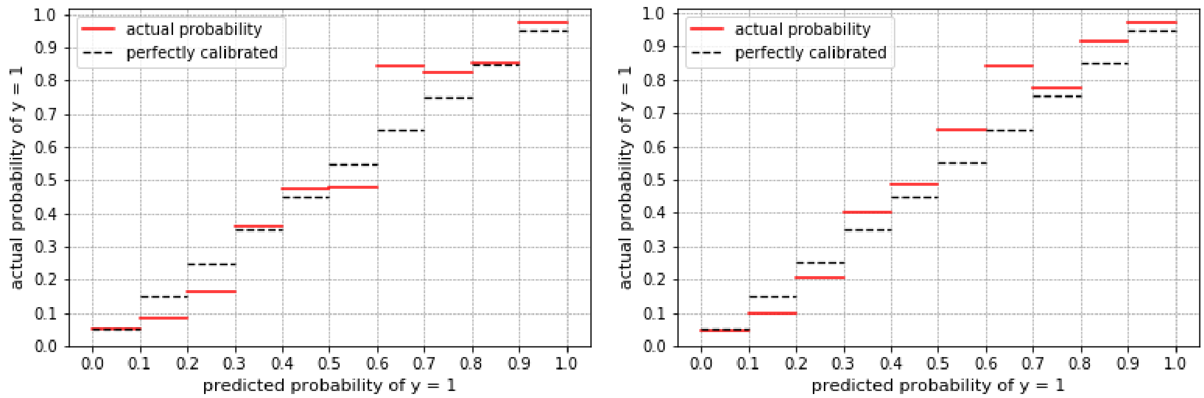

Figure 3: **Left**: $b = 0$. **Right**: learning b.

4 Isotonic regression

Figure 4 shows the reliability diagrams of Isotonic regression. It is worse than Platt scaling but is better than Squashing function.

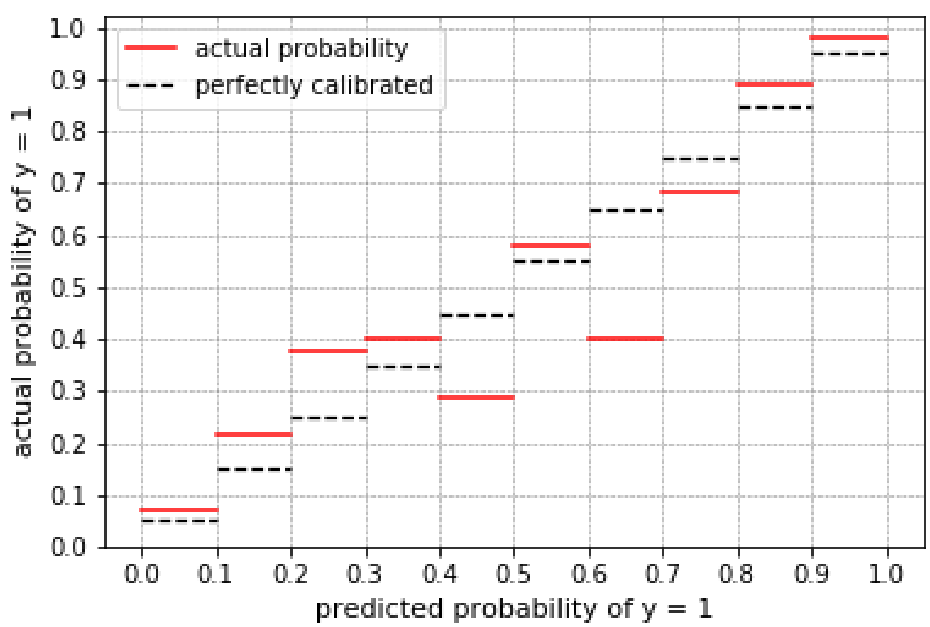

Figure 4: error rate versus different values of C

5 Discussion

Platt scaling is much simpler and is more effective when the predicted probability is sigmoid-shaped. Isotonic regression is a more complex and powerful calibration method to correct monotonic distortion. I would not always choose one over the other. I would choose Isotonic regression if the dataset is large enough since Isotonic regression is more prone to overfitting and would choose Platt scaling if dataset is scarce.

For multiclass setting, the calibration means the estimation of $P(y_j = 1|x)$ where j means the jth class. One way to obtain calibrated probability is to fit a 1 vs rest class of classifier and then plug each of those scores into the multinomial logistic regression instead of logistic regression.