# Importance Sampling Tree for Large-scale Empirical Expectation (supplementary material)

**Olivier Canévet** OLIVIER.CANEVET@IDIAP.CH
**Cijo Jose** CIJO.JOSE@IDIAP.CH
**François Fleuret** FRANCOIS.FLEURET@IDIAP.CH

## 1. An illustration of the `close` policy

Figure 1 illustrates the `close` policy. When the bounds of child 0 and child 1 of a node overlap, the probability of bifurcation is $0.5$. As soon as they no longer overlap, the score of the child with higher bounds is its lower one, and the score of the child with smaller bounds is its upper bound.

This policy explores more the tree and this is why it is slower in the end.

| $C$ | $\gamma$ | $T$ | Kernels | #SV | Acc. | Time (s) |
|---|---|---|---|---|---|---|
| 0.1 | 50 | 100000 | 265.24 | 7981 | 83.95 | 130.08078 |
| 0.5 | 50 | 100000 | 229.703 | 6213 | 83.95 | 125.11008 |
| 0.5 | 15 | 100000 | 357.499 | 4480 | 84.15 | 122.91283 |
| 0.5 | 10 | 100000 | 437.615 | 4185 | 84.25 | 122.89537 |
| 1 | 5 | 100000 | 620.246 | 3691 | 84.45 | 122.06775 |
| 5 | 5 | 100000 | 559.196 | 3217 | 83.95 | 120.50157 |
| 0.1 | 20 | 20000 | 369.676 | 7220 | 84.55 | 26.14553 |
| 0.1 | 5 | 20000 | 766.475 | 5758 | **84.75** | 25.972519 |
| 0.1 | 10 | 20000 | 513.464 | 6203 | **84.75** | 25.817479 |
| 0.5 | 50 | 20000 | **214.318** | 6213 | 83.95 | 25.072426 |
| 0.5 | 10 | 20000 | 396.372 | 4185 | 84.25 | 24.709539 |
| 0.1 | 50 | 10000 | 238.635 | 7981 | 83.95 | 13.116903 |
| 0.5 | 20 | 10000 | 279.822 | 4788 | 83.95 | 12.406951 |
| 0.1 | 5 | 5000 | 673.641 | 5758 | 84.15 | 6.62012 |
| 0.1 | 20 | 5000 | 337.403 | 7220 | 84.55 | 6.594196 |
| 0.1 | 10 | 5000 | 459.707 | 6203 | **84.75** | 6.57555 |
| 2 | 20 | 5000 | 235.904 | 3836 | 83.95 | 6.076082 |
| 0.1 | 5 | 2000 | 610.77 | 5758 | 84.15 | 2.746622 |
| 0.1 | 20 | 2000 | 314.86 | 7220 | 84.55 | 2.691372 |
| 0.1 | 15 | 2000 | 353.546 | 6766 | **84.75** | 2.674636 |
| 0.1 | 5 | 1000 | 562.106 | 5758 | **84.75** | 1.451517 |
| 0.1 | 10 | 1000 | 395.427 | 6203 | 84.25 | 1.392765 |
| 0.1 | 20 | 1000 | 298.203 | 7220 | 84.35 | 1.377687 |
| 0.5 | 15 | 1000 | 264.911 | 4480 | 84.05 | 1.30869 |
| 0.1 | 20 | 500 | 280.077 | 7220 | 84.15 | 0.736825 |
| 0.1 | 15 | 500 | 310.606 | 6766 | 84.45 | 0.729109 |
| 0.1 | 10 | 200 | 326.765 | 6203 | 84.65 | 0.347478 |
| 0.1 | 15 | 200 | 281.025 | 6766 | 83.95 | 0.334733 |
| 0.1 | 20 | 100 | 234.961 | 7220 | **84.75** | 0.197346 |
| 0.1 | 15 | 50 | 227.658 | 6766 | 84.15 | **0.128896** |

*Table 1.* Cross validation to choose parameters $C$ and $\gamma$. $T$ can be reduced for speed-up or increased for accuracy.

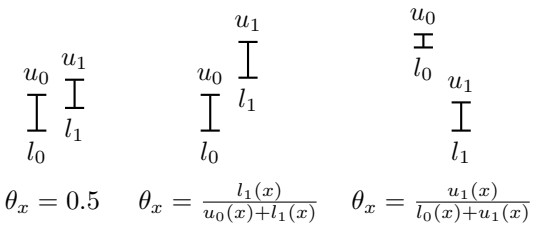


*Figure 1.* Illustration of the `close` policy

## 2. Checkerboard synthetic problem

The checkerboard data-set 2 is a 2-class problem made of $2D$-Gaussian clouds of points centred at each integer coordinate (from 0 to 10) and of standard deviation $0.3$ alternating classes.

Three parameters are involved in the IST SVM evaluation: $C$ and $\gamma$, the usual SVM parameters and $T$, the number of samplings with replacement that are performed in the evaluation. We chose these parameters through cross validation such that the validation accuracy and the actual number of kernel computation are Pareto optimal (see Table 1).

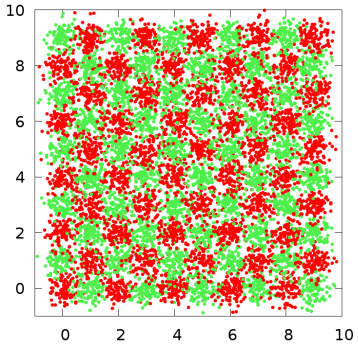


*Figure 2.* The $10 \times 10$ checkerboard toy problem – It is made of $2D$-Gaussian clouds centered at each integer coordinate and of standard deviation $0.3$ alternating classes.