A SUPPLEMENTARY MATERIAL

All the hyperparameters were set via a 5-fold cross-validation procedure. We present the details of our search space for each model.

LOW DIMENSIONAL REGIME

The coefficient of the error term C in LSVM, ℓ_2-regularized LR, and RSVM was obtained from the set $\{0.1, 1, 10, 100\}$. In the case of RSVM, we also chose the best kernel between a radial basis function (RBF), polynomials of degree 2 and 3, and sigmoid by performing a joint search over the kernels and C. The number of base predictors in the ensemble methods (RF, AB, GB), was searched over the set $\{10, 20, 50\}$. Both the square root and the log selection heuristics were explored to choose the maximum number of features in the RF estimators. The number of neighbors in kNN was optimized over the set $\{1, 3, 5, 7\}$. We considered the standard criteria, Gini index and entropy reduction, for attribute selection in DT. The candidate kernels in GP comprised of the RBF kernel scaled with scaled by a coefficient in the set $\{0.1, 1.0, 5\}$ and the dot product kernel with inhomogeneity 1. Finally, for SMP, λ and σ were fixed at 1, and τ at $0.01/p$.

HIGH DIMENSIONAL REGIME

The ℓ_1 regularization coefficient for each baseline was chosen from the set $\{0.1, 0.01, 0.001, 0.0001\}$. Moreover, in case of elastic net, the ratio of the ℓ_1 coefficient to the ℓ_2 coefficient was set to 1. For SMP, we fixed $p = 2$, $\lambda = 0.1$, $\eta = 0.01$ and $\tau = 0.01/p$, and searched for μ in the set $\{0.6, 0.7, 0.8, 0.9, 1\}$. All the results were averaged over 5 independent train-test splits of each dataset. We invoked the SGDClassifier implementation of the *scikit-learn* library for these baselines, and used the default settings for all the other hyperparameters.