A Bounds for kernel-based hypotheses

For the family of kernel-based hypotheses we consider for our algorithm, Corollary 1 leads to the following result.

Corollary 2 (Bounds for kernel-based hypotheses) *Let K be a PSD kernel, $\mathbb{H}$ the reproducing kernel Hilbert space associated to K and H the family of hypotheses defined by $H = \{h \in \mathbb{H} : \|h\|_K \leq \Lambda\}$, where $\Lambda \geq 0$. Assume that $K(x, x) \leq r$ for all $x \in \mathcal{X}$ for some $r > 0$. Fix $\tau \in [0, 1]$ and $\rho > 0$, then, for any $\delta > 0$, with probability at least $1 - \delta$ over the choice of a sample S of size m, each of the following holds for all $h \in H$:*

$$R(h, q_h) \leq \widehat{R}_\rho(h, q_h) + \frac{4r}{\rho} \frac{\Lambda}{\sqrt{m}} + \sqrt{\frac{\log \frac{1}{\delta}}{2m}}$$

$$R(h, q_h) \leq \widehat{R}_\rho(h, q_h) + \frac{r}{\rho} \frac{\Lambda}{\sqrt{m}} \left[\sqrt{\frac{\mathrm{Tr}[\mathbf{K}_{S_1}]}{mr^2}} + \sqrt{\frac{\mathrm{Tr}[\mathbf{K}_{S_2}]}{mr^2}} + 2 \right] + 3\sqrt{\frac{\log \frac{2}{\delta}}{2m}}.$$

Proof. By the reproducing property, for all $x \in \mathcal{X}$ and $h \in H$, $h(x) = \langle h, K(x, \cdot) \rangle$, thus $|h(x)| \leq \|h\|_K K(x, x) \leq \Lambda r$. The result follows using Theorem 1 and the following known upper bounds on the Rademacher complexity of H: $\widehat{\mathfrak{R}}_S(H) \leq \frac{\Lambda \sqrt{\mathrm{Tr}[\mathbf{K}]}}{m} \leq \sqrt{\frac{r^2 \Lambda^2}{m}}$, where $\mathbf{K}$ is the kernel matrix associated to sample S. $\square$

B Performance metrics

To evaluate and compare the AATP, INFINITEPUSH, and SVMPERF algorithms, we use the following metrics:

- Precision at τ (P@τ);
- Average precision (AP);
- Number of positives at the absolute top (Positives@top);
- Discounted cumulative gain at τ (DCG@τ);
- Normalized discounted cumulative gain at τ (NDCG@τ);

In all the definitions below, we assume that the items are enumerated according to the decreasing order of their scores. Also, for $i = 1, \ldots, N$, we denote by $\mathrm{rel}(i)$ the relevance of item i: $\mathrm{rel}(i) = 0$ if the item is negatively labeled, $\mathrm{rel}(i) = 1$ otherwise. We use P@quantile(i), $i = 1, \ldots, N$, to denote the percent quantile corresponding to top i items, i.e., P@quantile$(i) = i/N * 100$ (%).

Precision at the top (P@τ) equals the proportion of positive (relevant or preferred) instances among the top τ-quantile of score values, i.e.,

$$\text{P@}\tau = \frac{\text{\# of positive instances with score in top } \tau \text{ quantile}}{\text{\# of top } \tau\text{th quantile instances}}.$$

Average Precision (AP) equals to the average of the P@τ for all relevant items. Therefore,

$$\text{AP} = \frac{\sum_{i=1}^{N} \text{P@quantile}(i)\,\mathrm{rel}(i)}{\text{\# of positive instances}}.$$

Discounted Cumulative Gain (DCG@τ) is defined as

$$\text{DCG@}\tau = \sum_{i=1}^{\lceil \tau N \rceil} \frac{\mathrm{rel}(i)}{\log_2(i + 1)}.$$

Normalized Discounted Cumulative Gain (NDCG@τ) is calculated as

$$\text{NDCG@}\tau = Z_{\tau N} \sum_{i=1}^{\lceil \tau N \rceil} \frac{2^{\mathrm{rel}(i)} - 1}{\log(1 + i)}, \text{ with } Z_{\tau N} = \sum_{i=1}^{\lceil \tau N \rceil} \frac{1}{\log(1 + i)}.$$

Table 3: Spambase data: for each quantile value τ and each evaluation metric, there are two rows corresponding to the AATP (top) and the SVMPERF (bottom) algorithms.

τ (%)	P@τ	AP	DCG@τ	NDCG@τ	Positives@top
9	0.94 ± 0.05	0.9 ± 0.08	189.44 ± 2.80	0.94 ± 0.06	19.51 ± 30.00
	0.93 ± 0.02	0.88 ± 0.03	189.36 ± 1.10	0.94 ± 0.02	26.70 ± 33.77
6.5	0.93 ± 0.03	0.89 ± 0.02	189.54 ± 0.97	0.93 ± 0.03	18.11 ± 35.41
	0.94 ± 0.04	0.89 ± 0.03	189.35 ± 1.55	0.93 ± 0.04	18.55 ± 32.18
4.3	0.94 ± 0.04	0.84 ± 0.03	188.25 ± 1.03	0.95 ± 0.03	13.45 ± 17.53
	0.96 ± 0.01	0.88 ± 0.03	189.35 ± 1.07	0.96 ± 0.01	34.60 ± 36.73
2.2	0.97 ± 0.03	0.84 ± 0.03	188.25 ± 1.03	0.97 ± 0.02	13.5 ± 17.5
	0.99 ± 0.01	0.77 ± 0.02	185.80 ± 0.95	0.98 ± 0.01	3.61 ± 1.60
0.4	0.95 ± 0.01	0.84 ± 0.03	188.25 ± 1.03	0.96 ± 0.01	13.45 ± 17.53
	0.96 ± 0.03	0.84 ± 0.03	188.09 ± 0.88	0.96 ± 0.03	25.25 ± 38.14

C   Experimental results on Spambase data

The Spambase data set can be downloaded from the UCI Machine Learning Repository `http://archive.ics.uci.edu/ml/datasets.html`. The instances (4601 of them) correspond to email messages out of which 1813 are spam. The instances contain 57 features representing various word frequencies and other attributes. We split the data set into 20 independent subsets $S_1, S_2, \ldots, S_{20}$ each containing 5% (= 230) instances, except from S_{20} which contains 231 instances. Then, we ran 20 experiments, where in each experimental run, we used one of the subsets as the training set and merged all others to form the test set.

The data set is rather easy and the two algorithms obtain very comparable results.