A Proofs

A key concept to derive oracle inequalities and learning rates, which is used in the proof of Theorem 3.1, is the concept of entropy numbers, see Carl and Stephani (1990) or Steinwart and Christmann (2008, Definition A.5.26). Recall that, for normed spaces $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ as well as an integer $i \geq 1$, the i-th (dyadic) entropy number of a bounded, linear operator $S : E \to F$ is defined by

$$\begin{aligned} e_i(S : E \to F) &:= e_i(SB_E, \|\cdot\|_F) \\ &:= \inf\left\{\varepsilon > 0 : \exists s_1, \ldots, s_{2^{i-1}} \in SB_E \text{ such that } SB_E \subset \bigcup_{j=1}^{2^{i-1}} (s_j + \varepsilon B_F)\right\}, \end{aligned}$$

where we use the convention $\inf \emptyset := \infty$, and B_E as well as B_F denote the closed unit balls in E and F, respectively.

Proof of Theorem 3.1. We denote by $\tilde{H}$ the RKHS over X with Gaussian kernel of width $\gamma_{\max}$. Let $f_0 \in \tilde{H}$. Then, we can w.l.o.g. assume that $\|f_0\|_\infty \leq 1$, since the Gaussian kernel is bounded. For every $j \in \{1, \ldots m\}$ we define $f_j = \mathbf{1}_{A_j} f_0 = \widehat{f_{0|A_j}}$ and remark that $f_{j|A_j} \in H_{\gamma_{\max}}(A_j)$ due to Steinwart and Christmann (2008, Exercise 4.4i)). Hence, $f_j \in \hat{H}_{\gamma_{\max}}$ by definition of $\hat{H}_{\gamma_{\max}}$. Furthermore, since $\gamma_j \leq \gamma_{\max}$ for every $j \in \{1, \ldots m\}$, Steinwart and Christmann (2008, Proposition 4.4.6) shows that $\hat{H}_{\gamma_{\max}} \subset \hat{H}_{\gamma_j}$ with

$$\|f_j\|_{\hat{H}_{\gamma_j}} \leq \left(\frac{\gamma_{\max}}{\gamma_j}\right)^{d/2} \|f_j\|_{\hat{H}_{\gamma_{\max}}}. \tag{5}$$

Hence, we find that $f_j \in \hat{H}_{\gamma_j}$ for every $j \in \{1, \ldots, m\}$. Since

$$f_0 = \sum_{i=1}^m f_j$$

we conclude that $f_0 \in H$ by definition of H. Next, we observe with (5) that

$$\sum_{j=1}^m \lambda_j \|\mathbf{1}_{A_j} f_0\|^2_{\hat{H}_{\gamma_j}} = \sum_{j=1}^m \lambda_j \|f_j\|^2_{\hat{H}_{\gamma_j}} \leq \sum_{j=1}^m \lambda_j \left(\frac{\gamma_{\max}}{\gamma_j}\right)^d \|f_j\|^2_{\tilde{H}} \leq \sum_{j=1}^m \lambda_j \left(\frac{\gamma_{\max}}{\gamma_j}\right)^d \|f_0\|^2_{\tilde{H}}.$$

By using the latter inequality and the bound for the approximation error given in Steinwart and Christmann (2008, Theorem 8.18) with tail exponent $\tau = \infty$ since X is compact and with $\lambda = \sum_{j=1}^m \lambda_j \left(\frac{\gamma_{\max}}{\gamma_j}\right)^d$, we find that

$$\begin{aligned} \sum_{j=1}^m \lambda_j \|\mathbf{1}_{A_j} f_0\|^2_{\hat{H}_{\gamma_j}} + \mathcal{R}_{L,P}(f_0) - \mathcal{R}^*_{L,P} &\leq \sum_{j=1}^m \lambda_j \left(\frac{\gamma_{\max}}{\gamma_j}\right)^d \|f_0\|^2_{\tilde{H}} + \mathcal{R}_{L,P}(f_0) - \mathcal{R}^*_{L,P} \\ &\leq \max\{c_d, \tilde{c}_{d,\beta} c_{\mathrm{NE}}\} \left(\sum_{j=1}^m \lambda_j \left(\frac{\gamma_{\max}}{\gamma_j}\right)^d \gamma_{\max}^{-d} + \gamma_{\max}^\beta\right) \\ &\leq \hat{c}\left(\sum_{j=1}^m \lambda_j \gamma_j^{-d} + \gamma_{\max}^\beta\right), \end{aligned} \tag{6}$$

where $\hat{c} := \max\{c_d, \tilde{c}_{d,\beta} c_{\mathrm{NE}}\}$ with $c_d, \tilde{c}_{d,\beta} > 0$. Next, Eberts and Steinwart (2015, Theorem 6) provides the bound $e_i(\mathrm{id} : H_\gamma(A_j) \to L_2(\mathrm{P}_{X|A_j})) \leq a_j i^{-\frac{1}{2p}}$ for $i \geq 1$ with $a_j = \tilde{c}_p \sqrt{P_X(A_j)}\, r^{\frac{d+2p}{2p}} \gamma_j^{-\frac{d+2p}{2p}}$, where $\tilde{c}_p$ is a positive constant depending from p. For the constant a from Theorem B.1 this yields

$$\left(\max\left\{c_p m^{\frac{1}{2}} \left(\sum_{j=1}^m \lambda_j^{-p} a_j^{2p}\right)^{\frac{1}{2p}}, 2\right\}\right)^{2p}$$

$$
\begin{aligned}
&= \left(\max\left\{ c_p m^{\frac{1}{2}} \left(\sum_{j=1}^m \lambda_j^{-p} \left(\tilde{c}_p \sqrt{P_X(A_j)}\, r^{\frac{d+2p}{2p}} \gamma_j^{-\frac{d+2p}{2p}} \right)^{2p} \right)^{\frac{1}{2p}}, 2 \right\} \right)^{2p} \\
&= \left(\max\left\{ c_p \tilde{c}_p m^{\frac{1}{2}} r^{\frac{d+2p}{2p}} \left(\sum_{j=1}^m \left(\lambda_j^{-1} \gamma_j^{-\frac{d+2p}{p}} P_X(A_j) \right)^{p} \right)^{\frac{1}{2p}}, 2 \right\} \right)^{2p} \\
&\le \left(\max\left\{ c_p \tilde{c}_p m^{\frac{1}{2p}} r^{\frac{d+2p}{2p}} \left(\sum_{j=1}^m \lambda_j^{-1} \gamma_j^{-\frac{d+2p}{p}} P_X(A_j) \right)^{\frac{1}{2}}, 2 \right\} \right)^{2p} \\
&\le \left(\max\left\{ c_p \tilde{c}_p 16^{\frac{d}{2p}} r \left(\sum_{j=1}^m \lambda_j^{-1} \gamma_j^{-\frac{d+2p}{p}} P_X(A_j) \right)^{\frac{1}{2}}, 2 \right\} \right)^{2p} \\
&\le C_p r^{2p} \left(\sum_{j=1}^m \lambda_j^{-1} \gamma_j^{-\frac{d+2p}{p}} P_X(A_j) \right)^{p} + 4^p \\
&=: a^{2p} ,
\end{aligned}
$$

where we used that $\|\cdot\|_p \le m^{\frac{1-p}{p}} \|\cdot\|_1$ for $0 < p < 1$, as well as $mr^d \le 16^d$ by (3) and that $C_p := c_p^{2p} \tilde{c}_p^{2p} 16^d$. Then, by using Theorem B.1, (6), the concavity of the function $t \mapsto t^{\frac{q+1}{q+2-p}}$ for $t \ge 0$ and the fact that $\tau \ge 1$ with $\tau \le n$ we obtain that

$$
\begin{aligned}
&\sum_{j=1}^m \lambda_j \|f_{\mathrm{D}_j,\lambda_j,\gamma_j}\|_{\hat{H}_j}^2 + \mathcal{R}_{L,P}(\widehat{f}_{\mathrm{D},\boldsymbol{\lambda},\boldsymbol{\gamma}}) - \mathcal{R}_{L,P}^* \\
&\le 9 \left(\sum_{j=1}^m \lambda_j \|\mathbf{1}_{A_j} f_0\|_{\hat{H}_{\gamma_j}}^2 + \mathcal{R}_{L,P}(f_0) - \mathcal{R}_{L,P}^* \right) + C \left(a^{2p} n^{-1} \right)^{\frac{q+1}{q+2-p}} + 3 \left(\frac{432 c_{\mathrm{NE}}^{\frac{q}{q+1}} \tau}{n} \right)^{\frac{q+1}{q+2}} + \frac{30\tau}{n} \\
&\le 9\hat{c} \left(\sum_{j=1}^m \lambda_j \gamma_j^{-d} + \gamma_{\max}^{\beta} \right) + C \left(C_p r^{2p} \left(\sum_{j=1}^m \lambda_j^{-1} \gamma_j^{-\frac{d+2p}{p}} P_X(A_j) \right)^{p} n^{-1} + 4^p n^{-1} \right)^{\frac{q+1}{q+2-p}} \\
&\quad + 3 \left(\frac{432 c_{\mathrm{NE}}^{\frac{q}{q+1}} \tau}{n} \right)^{\frac{q+1}{q+2}} + \frac{30\tau}{n} \\
&\le 9\hat{c} \left(\sum_{j=1}^m \lambda_j \gamma_j^{-d} + \gamma_{\max}^{\beta} \right) + C \left(C_p r^{2p} \left(\sum_{j=1}^m \lambda_j^{-1} \gamma_j^{-\frac{d+2p}{p}} P_X(A_j) \right)^{p} n^{-1} \right)^{\frac{q+1}{q+2-p}} \\
&\quad + C \left(\frac{4^p \tau}{n} \right)^{\frac{q+1}{q+2-p}} + 3 \left(\frac{432 c_{\mathrm{NE}}^{\frac{q}{q+1}} \tau}{n} \right)^{\frac{q+1}{q+2}} + \frac{30\tau}{n} \\
&\le \tilde{C}_{\beta,d,p,q} \left(\sum_{j=1}^m \lambda_j \gamma_j^{-d} + \gamma_{\max}^{\beta} + \left(r^{2p} \left(\sum_{j=1}^m \lambda_j^{-1} \gamma_j^{-\frac{d+2p}{p}} P_X(A_j) \right)^{p} n^{-1} \right)^{\frac{q+1}{q+2-p}} \right. \\
&\quad \left. + \left(\frac{\tau}{n} \right)^{\frac{q+1}{q+2-p}} + \left(\frac{\tau}{n} \right)^{\frac{q+1}{q+2}} + \frac{\tau}{n} \right) \\
&\le C_{\beta,d,p,q} \left(\sum_{j=1}^m \lambda_j \gamma_j^{-d} + \gamma_{\max}^{\beta} + \left(r^{2p} \left(\sum_{j=1}^m \lambda_j^{-1} \gamma_j^{-\frac{d+2p}{p}} P_X(A_j) \right)^{p} n^{-1} \right)^{\frac{q+1}{q+2-p}} + \left(\frac{\tau}{n} \right)^{\frac{q+1}{q+2}} \right)
\end{aligned}
$$

holds with probability P^n not less than $1 - 3e^{-\tau}$, where the constants $\tilde{C}_{\beta,d,p,q}$ and $C_{\beta,d,p,q}$ are given by

$\tilde{C}_{\beta,d,p,q} := \max\{9\hat{c}, C\cdot C_p^{\frac{q+1}{q+2-p}}, C\cdot 4^{\frac{p(q+1)}{q+2-p}}, 3c_{\mathrm{NE}}^{\frac{q}{q+2}}\cdot 432^{\frac{q+1}{q+2}}, 30\}$ and $C_{\beta,d,p,q} := \max\{9\hat{c}, C\cdot C_p^{\frac{q+1}{q+2-p}}, 3\cdot C\cdot 4^{\frac{p(q+1)}{q+2-p}}, 9c_{\mathrm{NE}}^{\frac{q}{q+2}}\cdot 432^{\frac{q+1}{q+2}}, 90\}$. □

Proof of Theorem 3.2. First we simplify the presentation by using the sequences $\tilde{\lambda}_n := c_2 n^{-(\beta+d)\kappa}$ and $\tilde{\gamma}_n := c_3 n^{-\kappa}$ with $\kappa := \frac{(q+1)}{\beta(q+2)+d(q+1)}$. Then, we find with Theorem 3.1 together with $r_n = c_1 n^{-\nu}, \lambda_{n,j} = r_n^d \tilde{\lambda}_n$ and $\gamma_{n,j} = \tilde{\gamma}_n$ and with $\sum_{j=1}^{m_n} P_X(A_j) = 1$ and $m_n \leq 16^d r_n^{-d}$ that

$$
\begin{aligned}
&\mathcal{R}_{L,P}(\widehat{f}_{\mathrm{D},\boldsymbol{\lambda},\boldsymbol{\gamma}}) - \mathcal{R}_{L,P}^* \\
&\leq C_{\beta,d,p,q}\left(\sum_{j=1}^{m_n} \lambda_{n,j}\gamma_{n,j}^{-d} + \gamma_{\max}^{\beta} + \left(r_n^{2p}\left(\sum_{j=1}^{m_n} \lambda_{n,j}^{-1}\gamma_{n,j}^{-\frac{d+2p}{p}} P_X(A_j)\right)^p n^{-1}\right)^{\frac{q+1}{q+2-p}} + \left(\frac{\tau}{n}\right)^{\frac{q+1}{q+2}}\right) \\
&= C_{\beta,d,p,q}\left(m_n r_n^d \tilde{\lambda}_n \tilde{\gamma}_n^{-d} + \tilde{\gamma}_n^{\beta} + \left(r_n^{2p}\left(r_n^{-d}\tilde{\lambda}_n^{-1}\tilde{\gamma}_n^{-\frac{d+2p}{p}} \sum_{j=1}^{m_n} P_X(A_j)\right)^p n^{-1}\right)^{\frac{q+1}{q+2-p}} + \left(\frac{\tau}{n}\right)^{\frac{q+1}{q+2}}\right) \\
&\leq 16^d C_{\beta,d,p,q}\left(\tilde{\lambda}_n\tilde{\gamma}_n^{-d} + \tilde{\gamma}_n^{\beta} + \left(r_n^{p(2-d)}\tilde{\lambda}_n^{-p}\tilde{\gamma}_n^{-(d+2p)} n^{-1}\right)^{\frac{q+1}{q+2-p}} + \left(\frac{\tau}{n}\right)^{\frac{q+1}{q+2}}\right) \\
&\leq C_{\beta,\nu,d,p,q}\left(n^{-(\beta+d)\kappa} n^{+d\kappa} + n^{-\beta\kappa} + \left(\frac{n^{-p\nu(2-d)}}{n^{-p(\beta+d)\kappa} n^{-(d+2p)\kappa+1}}\right)^{\frac{q+1}{q+2-p}} + \left(\frac{\tau}{n}\right)^{\frac{q+1}{q+2}}\right) \\
&= C_{\beta,\nu,d,p,q}\left(2n^{-\beta\kappa} + \left(\frac{n^{-p[\nu(2-d)-(\beta+d)\kappa-2\kappa]}}{n^{-d\kappa+1}}\right)^{\frac{q+1}{q+2-p}} + \left(\frac{\tau}{n}\right)^{\frac{q+1}{q+2}}\right) \\
&= C_{\beta,\nu,d,p,q}\left(2n^{-\beta\kappa} + \frac{n^{-\frac{p(q+1)}{q+2-p}[\nu(2-d)-(\beta+d+2)\kappa]}}{n^{\frac{\beta(q+2)(q+1)}{(\beta(q+2)+d(q+1))(q+2-p)}}} + \left(\frac{\tau}{n}\right)^{\frac{q+1}{q+2}}\right) \\
&\leq C_{\beta,\nu,d,p,q}\left(2n^{-\beta\kappa} + \frac{n^{-\frac{p(q+1)}{q+2-p}[\nu(2-d)-(\beta+d+2)\kappa]}}{n^{\frac{\beta(q+1)}{(\beta(q+2)+d(q+1))}}} + \left(\frac{\tau}{n}\right)^{\frac{q+1}{q+2}}\right) \\
&\leq C_{\beta,\nu,d,p,q}\left(2n^{-\frac{\beta(q+1)}{\beta(q+2)+d(q+1)}} + n^{-\frac{\beta(q+1)}{\beta(q+2)+d(q+1)}} n^{-\frac{p(q+1)}{q+2}\left[\nu(2-d)-\frac{(\beta+d+2)(q+1)}{\beta(q+2)+d(q+1)}\right]} + \left(\tau n^{-1}\right)^{\frac{q+1}{q+2}}\right) \\
&\leq C_{\beta,\nu,\xi,d,q}\tau^{\frac{q+1}{q+2}} \cdot n^{-\frac{\beta(q+1)}{\beta(q+2)+d(q+1)}+\xi}
\end{aligned}
$$

holds with probability P^n not less than $1-3e^{-\tau}$, where the constants $C_{\beta,\nu,d,p,q}, c_1, c_2, c_3 > 0$ depend on β, ν, d, p, q and the constant $C_{\beta,\nu,\xi,d,q} > 0$ depends on β, ν, ξ, d, q. Furthermore, we remark that we chose p sufficiently close to zero such that $\xi \geq \frac{p(q+1)}{q+2}\left(\nu(d-2) + \frac{(\beta+d+2)(q+1)}{\beta(q+2)+d(q+1)}\right) \geq 0$. □

B Appendix

Theorem B.1. *Let P be a distribution on $X \times Y$ with noise exponent $q \in (0,\infty]$ and let $L : Y \times \mathbb{R} \to [0,\infty]$ be the hinge loss. Furthermore let (A) be satisfied with $H_j := H_{\gamma_j}(A_j)$ and assume that, for fixed $n \geq 1$, there exist constants $p \in (0,1)$ and $a_1, \ldots, a_m > 0$ such that for all $j \in \{1, \ldots, m\}$*

$$
e_i(\mathrm{id} : H_j \to L_2(\mathrm{P}_{X|A_j})) \leq a_j\, i^{-\frac{1}{2p}}, \qquad i \geq 1. \tag{7}
$$

Finally, fix an $f_0 \in H$. Then, for all fixed $\tau > 0$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m) > 0$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m) > 0$ and

$$
a := \max\left\{c_p m^{\frac{1}{2}}\left(\sum_{j=1}^{m} \lambda_j^{-p} a_j^{2p}\right)^{\frac{1}{2p}}, 2\right\}
$$

the VP-SVM given by (4) satisfies

$$\sum_{j=1}^{m} \lambda_j \|f_{D_j,\lambda_j,\gamma_j}\|_{\hat{H}_j}^2 + \mathcal{R}_{L,P}(\widehat{f}_{D,\boldsymbol{\lambda},\boldsymbol{\gamma}}) - \mathcal{R}_{L,P}^*$$
$$\leq 9 \left(\sum_{j=1}^{m} \lambda_j \|\mathbf{1}_{A_j} f_0\|_{\hat{H}_j}^2 + \mathcal{R}_{L,P}(f_0) - \mathcal{R}_{L,P}^* \right) + C \left(a^{2p} n^{-1}\right)^{\frac{q+1}{q+2-p}} + 3 \left(\frac{432 c_{NE}^{\frac{q}{q+1}} \tau}{n} \right)^{\frac{q+1}{q+2}} + \frac{30\tau}{n}$$

with probability P^n *not less than* $1 - 3e^{-\tau}$*, where* $C > 0$ *is a constant only depending on* p*.*

Proof of Theorem B.1. One can obtain the result directly by an application of Eberts and Steinwart (2015, Theorem 5). To this end, we note that the hinge loss is Lipschitz continuous and can be clipped at $M = 1$. Since H is the sum-RKHS of RKHSs with Gaussian kernels and the Gaussian kernel is bounded, w.l.o.g. we assume for $f_0 \in H$ that $\|f_0\|_\infty \leq 1$. Hence, $\|L \circ f_0\|_\infty \leq 2$ and therefore $B_0 = 2$. Furthermore Steinwart and Christmann (2008, Theorem 8.24) showed that for the hinge loss the constants V and ϑ from Eberts and Steinwart (2015, Theorem 5) can be achieved by $V = 6c_{\mathrm{NE}}^{\frac{q}{q+1}}$ and $\vartheta = \frac{q}{q+1}$. That means

$$\sum_{j=1}^{m} \lambda_j \|f_{D_j,\lambda_j,\gamma_j}\|_{\hat{H}_j}^2 + \mathcal{R}_{L,P}(\widehat{f}_{D,\boldsymbol{\lambda},\boldsymbol{\gamma}}) - \mathcal{R}_{L,P}^*$$
$$\leq 9 \left(\sum_{j=1}^{m} \lambda_j \|\mathbf{1}_{A_j} f_0\|_{\hat{H}_j}^2 + \mathcal{R}_{L,P}(f_0) - \mathcal{R}_{L,P}^* \right) + C \left(\frac{a^{2p}}{n} \right)^{\frac{1}{2-p-\vartheta+\vartheta p}} + 3 \left(\frac{72 V \tau}{n} \right)^{\frac{1}{2-\vartheta}} + \frac{15 B_0 \tau}{n}$$
$$= 9 \left(\sum_{j=1}^{m} \lambda_j \|\mathbf{1}_{A_j} f_0\|_{\hat{H}_j}^2 + \mathcal{R}_{L,P}(f_0) - \mathcal{R}_{L,P}^* \right) + C \left(\frac{a^{2p}}{n} \right)^{\frac{q+1}{q+2-p}} + 3 \left(\frac{432 c_{\mathrm{NE}}^{\frac{q}{q+1}} \tau}{n} \right)^{\frac{q+1}{q+2}} + \frac{30\tau}{n}$$

holds with probability P^n not less than $1 - 3e^{-\tau}$. □

C Some more details for results

In this section we give some more technical details the pseudo-code for local SVMs (Algorithm 1), and some more results of the experiments. Firstly, some more details on how the experiments were performed:

Hyperparameter grid We used `liquidSVM`'s default grid: the λ are geometrically spaced between $0.01/\tilde{n}$ and $0.001/\tilde{n}$ where $\tilde{n}$ is the number of samples contained in the $k-1$ folds currently used for training. The γ are geometrically spaced between $5r$ and $0.2r\tilde{n}^{-1/d}$ where r is the radius of the cell, d is the dimension of the data and $\tilde{n}$ is as above.

Spatial partitioning scheme The segmentation for mid-sized data sets $n \leq 50000$ finds centers for the Voronoi cells using the farthest-first-traversal algorithm on the entire data set. For larger data sets a random subsample of the full data set is created and the splitting described above is applied recursively.

In `liquidSVM`, this is achieved with value `6` for the `partition` argument to `scripts/mc-svm.sh` (or the `-P 6` mode of `svm-train`).

Random Chunks scheme The data is split into random partitions of the specified size. In `liquidSVM`, this is achieved with value `1` for the `partition` argument to `scripts/mc-svm.sh` (or the `-P 1` mode of `svm-train`).

Algorithm 1 Local SVM training and testing.

Require: A training dataset D, split into cells $D_1, \dots, D_m$, $m \geq 1$, a set $\Gamma \subset \mathbb{R}_{>0}$ of γ-candidates, a set $\Lambda \subset \mathbb{R}_{>0}$ of λ-candidates, the number of folds k for cross-validation, and a test set D^T, split into cells $D_1^T, \dots, D_m^T$.

Ensure: Test error

for all $j = 1, \dots, m$ **do**
  Split the cell D_j into k random parts $D_{j,1}, \dots, D_{j,k}$.
  for all $\ell = 1, \dots, k$ **do**
    $D'_{j,\ell} := D_j \setminus D_{j,\ell}$
    cache pre-kernel matrix $(x_1, x_2) \to \|x_1 - x_2\|_2$ for x_1, x_2 in $D'_{j,\ell}$
    for all $\gamma \in \Gamma$ **do**
      use cached pre-kernel matrix to calculate kernel matrix with bandwidth γ
      for all $\lambda \in \Lambda$ **do**
        Train an SVM $f_{D'_{j,\ell},\lambda,\gamma}$ of the form (4) (possibly using as warm-start the solution for the previous λ-candidate).
        Calculate and save the validation risk $\mathcal{R}_{L,D_{j,\ell}}(f_{D'_{j,\ell},\lambda,\gamma})$
      end for
      Let $f_{D_j,\lambda,\gamma}$ be the linear combination of the $(f_{D'_{j,\ell},\lambda,\gamma})_{1\leq\ell\leq k}$ with weights exponential in $\mathcal{R}_{L,D_{j,\ell}}(f_{D'_{j,\ell},\lambda,\gamma})$.
      Save the validation risk $\mathcal{R}_{L,D_j}(f_{D_j,\lambda,\gamma})$
    end for
  end for
end for
for all $j = 1, \dots, m$ **do**
  Select the γ_j, λ_j-combination minimizing the combined validation risk.
end for
for all $j = 1, \dots, m$ **do**
  Calculate test error $\mathcal{R}_{L,D_j^T}(f_{D_j,\lambda_j,\gamma_j})$ on test cell D_j^T.
end for
return global test error $\frac{1}{|D^T|}\sum_{j=1}^m |D_j^T| \cdot \mathcal{R}_{L,D_j^T}(f_{D_j,\lambda_j,\gamma_j})$.

Table 6: Training times divided by the product of training set size, cell size and dimension (in seconds times 10^9). This shows that the time complexity in (1) is fulfilled quite nicely.

	2000	5000	10000	15000
	training time			
HIGGS	33	29	29	29
HEPMASS	22	19	19	20
GASSENSOR	20	19	19	19
SUSY	45	39	38	38
COVTYPE	10	9	9	9
COD-RNA	52	51	45	48
SKIN	116	117	106	111

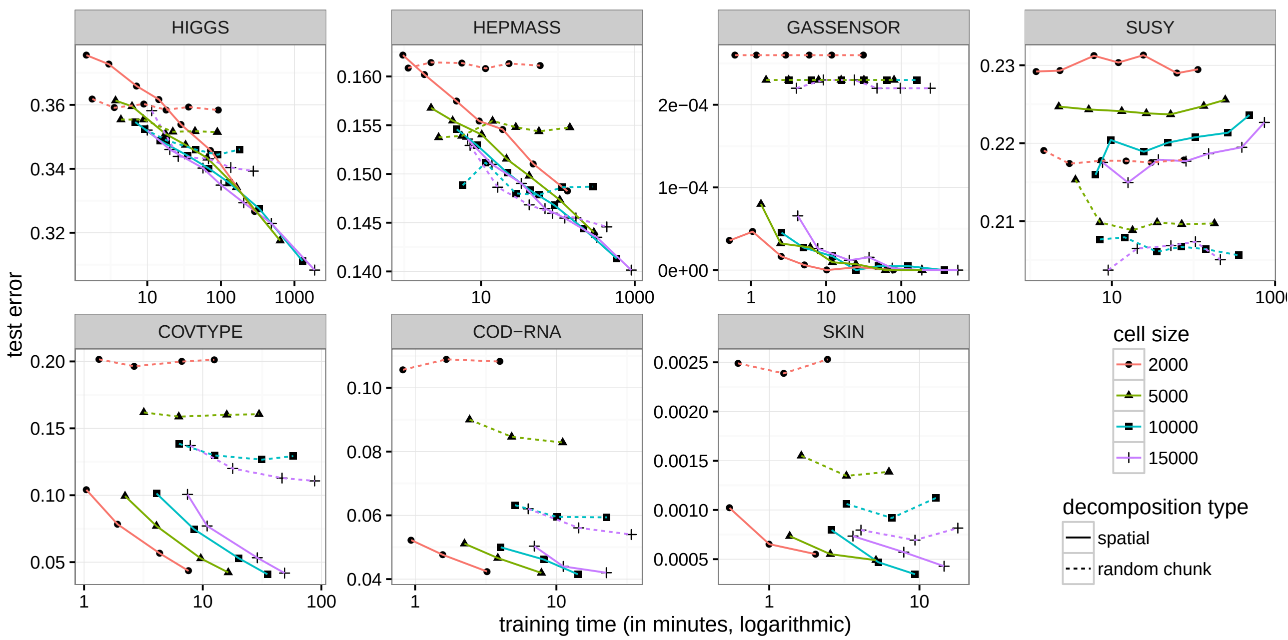

Figure 4: Training time vs. test error, this is the final trade-off.