Regularized Empirical Risk Minimization

Problem Setting:
Unknown distribution: rv Z ∈ Z with distribution ρ
Parameter θ ∈ H, H a Hilbert space
Problem: Minimize an expected loss:
\[ \min_{\theta} L(\theta) := \mathbb{E} \left[ \ell(\theta) \right], \quad \ell(\theta) \text{ loss function} \]

Well-specified assumption \( \theta^* \in \arg\min_{\theta \in H} L(\theta) \)
Statistical performance: \( L(\theta) - L(\theta^*) \)
Data: access to \( \rho \) through \( n \) i.i.d observations \( (z_i)_{i \leq n} \)
From Z

Regularized ERM:
\[ \hat{\theta}_\lambda = \arg\min_{\theta \in H} L(\theta) + \frac{\lambda}{2} \| \theta \|^2 \]
Basic result: slow rates
\[ L(\hat{\theta}_\lambda) - L(\theta^*) \leq \frac{\| \nabla L(\theta^*) \|^2}{\lambda n} \]

Bias-variance trade-off for least-squares
Loss: \( \ell(\theta \cdot x) = \| y - \theta : x \|^2 \)
Covariance operator: \( \Sigma = \mathbb{E} [ xx^T ] \)
\[ \forall \theta \in H, \ L(\theta) - L(\theta^*) = \| \Sigma^{1/2} (\theta - \theta^*) \|^2 = \| \theta - \theta^* \|^2 \Sigma \]
Two main terms:
Effective dimension: \( \text{df}_\lambda = \text{Tr}(\Sigma^{1/2} \Sigma^{1/2}) \Sigma \)
\[ \Sigma = \Sigma + \lambda I \]
Bias term:
\[ L(\hat{\theta}_\lambda) - L(\theta^*) \leq \frac{\| \nabla L(\theta^*) \|^2}{\lambda n} \]

Parametrization and optimal rates
Effective dimension ↔ spectrum of covariance matrix \( \Sigma \), eigenvalues of \( \Sigma \) in decreasing order.
Assumption: \( \text{df}_\lambda \leq Q^3 \lambda^{1/4} \)\( \lambda = O(i^{-n}) \)
Bias term ↔ difficulty of the learning problem assumption: \( \text{df}_\lambda \leq \lambda^{1/2} \) \( \Rightarrow \| \Sigma^{1/2} \theta - \Sigma^{1/2} \theta^* \|^2 \leq \gamma \text{Tr}(\Sigma) \)
Optimal fast rates for \( \lambda = (Q/L)^2 n^{-o((1+2\gamma)n+1)} \)
\[ L(\hat{\theta}_\lambda) - L(\theta^*) \leq \sqrt{\frac{\text{df}_\lambda}{\lambda}} \]

We acknowledge support from the ERCIM Alain Bensoussan Fellowship and the INRIA, Département d’Informatique de l’ENS, PSL Research University.

Acknowledgments and References