On Minimax Rates in Active Learning for Functions in the Rashomon Set

1 Introduction

Consider the problem of active learning of functions in a Hölder smooth class. Previous work has established minimax rates for this setting. However, in many practical scenarios, we may have additional information that restricts our attention to functions that perform similarly to some reference function - the so-called Rashomon set. This paper examines how such a restriction affects the minimax rates of convergence.

2 Setup and Definitions

Let $\Sigma(L, \alpha)$ denote the class of Hölder smooth functions $f : [0,1]^d \to \mathbb{R}$ with parameters $L, \alpha > 0$, such that for all $x, z \in [0,1]^d$:

$$|f(z) - P_x(z)| \leq L\|z - x\|^\alpha$$

where $P_x(z)$ is the order $k = \lfloor \alpha \rfloor$ Taylor polynomial of f around x.

Given a reference function $\hat{f}$ and parameter $\theta > 0$, we define the Rashomon set:

$$\hat{R}_{set}(F, \theta) = \{f \in \Sigma(L, \alpha) : \|f - \hat{f}\|^2 \leq \theta\}$$

Assumption 1. *The reference function $\hat{f}$ belongs to $\Sigma(L, \alpha)$.*

3 Main Results

Let $\delta(\theta, n)$ be the solution to:

$$\min\{c_1\delta^{-d/\alpha}, d\log(\sqrt{\theta}/\delta)\} = cn$$

where c_1, c are positive constants.

Lemma 1 (Intersection Point)**.** *For any $\theta > 0$, there exists a unique point $\delta^*(\theta)$ where:*

$$c_1(\delta^*)^{-d/\alpha} = d\log(\sqrt{\theta}/\delta^*)$$

Moreover, $\delta^*(\theta)$ satisfies:

$$\delta^*(\theta) = \theta^{\frac{\alpha}{2\alpha+d}}(\log(1/\theta))^{-\alpha/d}$$

Proof. Taking logarithms of both sides:

$$\log(c_1) - \frac{d}{\alpha}\log(\delta^*) = \log(d) + \log(\log(\sqrt{\theta}/\delta^*))$$

The left side is strictly decreasing in δ^* while the right side is strictly increasing in δ^* for $\delta^* < \sqrt{\theta}$. Therefore, there exists a unique solution. The form of $\delta^*(\theta)$ can be verified by substitution. $\square$

This intersection point determines three regimes:

Theorem 1 (Minimax Rates by Regime)**.** *Let $n \geq 1$ and $\theta > 0$. Then:*

$$\inf_{(\hat{f}_n, S_n)\in\Theta_{active}} \sup_{f\in\hat{R}_{set}(F,\theta)} \mathbb{E}\|\hat{f}_n - f\|^2 \geq R(n,\theta)$$

where $R(n,\theta)$ depends on the regime:

1. *Small θ regime ($\theta \leq n^{-2\alpha/d}$):*
$$R(n,\theta) = \theta$$

2. *Large θ regime ($\theta \geq n^{-2\alpha/(2\alpha+d)}$):*
$$R(n,\theta) = n^{-2\alpha/(2\alpha+d)}$$

3. *Intermediate regime:*
$$R(n,\theta) = \theta\exp(-2cn/d)$$

Proof. The proof proceeds by carefully analyzing how the constraints imposed by the Rashomon set interact with the packing number bounds required for the minimax lower bound. Let us first build intuition for the three regimes before proceeding with the technical details.

In the small θ regime, the Rashomon set is so constrained that all functions within it must be very close to $\hat{f}$. The maximum separation between any two functions is bounded by $\sqrt{2\theta}$, and this dominates any other constraints. No matter how many samples we take, we cannot improve beyond this intrinsic limitation of the Rashomon set.

In the large θ regime, the Rashomon set is sufficiently large that it does not meaningfully constrain the construction needed for the minimax lower bound. We can construct the same set of well-separated functions used in the classical proof, and these functions all lie within the Rashomon set. Thus, we recover the original minimax rate.

The intermediate regime is the most interesting - here the Rashomon set constraint and the sample size interact in a non-trivial way. The constraint limits how far apart we can separate our functions, but we can still construct enough well-separated functions to force a meaningful lower bound that depends on both n and θ.

Let us now proceed with the technical details. First, we establish a key lemma about packing numbers in the Rashomon set:

Lemma 2 (Packing Numbers)**.** *For any $\epsilon > 0$, let $M(\epsilon,\theta)$ be the maximum number of functions $\{g_1,\ldots,g_m\}$ such that:*

- $\|g_i - \hat{f}\|^2 \leq \theta$ *for all* i
- $\|g_i - g_j\|^2 \geq \epsilon$ *for all* $i \neq j$

- *Each* $g_i \in \Sigma(L, \alpha)$

Then:

$$M(\epsilon, \theta) \leq \min\{\exp(c_1 \epsilon^{-d/\alpha}), (\sqrt{\theta}/\epsilon)^d\}$$

Proof of Packing Numbers Lemma. The first term comes from the metric entropy of Hölder smooth functions. The second term comes from volume considerations - we are packing ϵ-balls into a ball of radius $\sqrt{\theta}$ in L^2 metric in d dimensions. □

Now for each regime, we construct an appropriate set of functions and apply Fano's inequality:

Small θ regime ($\theta \leq n^{-2\alpha/d}$): In this regime, consider any two functions $f_1, f_2 \in \hat{R}_{set}(F, \theta)$. By the triangle inequality:

$$\|f_1 - f_2\|^2 \leq 2\theta$$

Therefore, no estimator can achieve error better than θ regardless of sample size.

Large θ regime ($\theta \geq n^{-2\alpha/(2\alpha+d)}$): Here we can use the classical construction from Castro et al. Let $\{h_1, \ldots, h_M\}$ be their set of well-separated functions with separation $\epsilon = n^{-\alpha/(2\alpha+d)}$. Define:

$$f_i = \hat{f} + h_i - \mathbb{E}[h_i]$$

These functions maintain the required separation while staying within the Rashomon set due to the regime condition on θ. The classical proof then applies directly.

Intermediate regime: Here we must carefully balance the constraints. From the packing numbers lemma, we need:

$$\log M(\epsilon, \theta) \leq cn$$

This gives:

$$\min\{c_1 \epsilon^{-d/\alpha}, d \log(\sqrt{\theta}/\epsilon)\} = cn$$

Solving this equation in the intermediate regime yields:

$$\epsilon = \sqrt{\theta} \exp(-cn/d)$$

Therefore:

$$R(n, \theta) = \theta \exp(-2cn/d)$$

The regime transitions occur exactly where these rates meet:

- At $\theta = n^{-2\alpha/d}$, the small and intermediate regimes give the same rate
- At $\theta = n^{-2\alpha/(2\alpha+d)}$, the intermediate and large regimes give the same rate

□

[Previous sections remain the same through the proof, then adding:]

4 Technical Conditions and Transition Points

The transition points between regimes emerge naturally from the interaction between three fundamental constraints:

1. The Rashomon set constraint: $\|f - \hat{f}\|^2 \leq \theta$
2. The packing number constraint: We need enough well-separated functions to force a lower bound
3. The Hölder smoothness constraint: All constructed functions must remain in $\Sigma(L, \alpha)$

To make this precise, we need additional technical conditions:

Assumption 2 (Regularity of Construction)**.** *For any $f_1, f_2 \in \hat{R}_{set}(F, \theta)$, their difference $f_1 - f_2$ satisfies:*

$$\|D^k(f_1 - f_2)\|_\infty \leq 2L$$

for all multi-indices k with $|k| \leq \lfloor \alpha \rfloor$, where D^k denotes the kth partial derivative.

This condition ensures our constructed functions maintain Hölder smoothness. It is satisfied by standard constructions using B-spline bases.

The transition point $\theta = n^{-2\alpha/d}$ between the small and intermediate regimes occurs precisely when:

$$\theta = \theta \exp(-2cn/d)$$

This equality reflects where the pure Rashomon set constraint begins to dominate the statistical estimation error. When θ is smaller than this threshold, no amount of sampling can overcome the intrinsic limitation imposed by the small Rashomon set.

The transition point $\theta = n^{-2\alpha/(2\alpha+d)}$ between the intermediate and large regimes occurs when:

$$\theta \exp(-2cn/d) = n^{-2\alpha/(2\alpha+d)}$$

This is where the Rashomon set becomes large enough that it no longer meaningfully constrains the classical construction. Intuitively, this happens when the volume of the Rashomon set exceeds the volume needed to pack functions at the classical minimax separation rate.

5 Implications and Examples

The different regimes have important implications for active learning in practice:

Example 1 (High-Dimensional Smooth Function)**.** *Consider estimating a function $f : [0,1]^{10} \to \mathbb{R}$ with $\alpha = 2$ (twice differentiable). In the classical setting, the minimax rate is $n^{-4/14}$. However, if we know f lies in a Rashomon set with $\theta = n^{-1/2}$, we are in the small θ regime and achieve rate $\theta = n^{-1/2}$, substantially better than the classical rate.*

Example 2 (Image Boundary Detection)**.** *In the context of Korostelev's image boundary detection problem, the Rashomon set constraint effectively reduces the search space near the boundary. When θ is properly scaled with n, this can lead to improved rates in the intermediate regime without requiring the boundary to be exactly known.*

These examples highlight a broader principle: The Rashomon set constraint can be viewed as a form of regularization that becomes particularly powerful in high dimensions or when the function class has localized complexity.

6 Practical Implications

The existence of these three regimes has important implications for algorithm design:

1. In the small θ regime, one should focus on efficient exploration of the Rashomon set rather than sophisticated active learning strategies, as the constraint dominates the statistical error.

2. In the intermediate regime, active learning strategies should be adapted to account for the Rashomon set constraint. This suggests new algorithms that combine classical active learning criteria with constraints to stay within the Rashomon set.

3. In the large θ regime, classical active learning strategies remain optimal, but the Rashomon set can still provide computational benefits by restricting the search space.

Remark 1. *The rates we derive are minimax lower bounds. Achieving these rates with computationally efficient algorithms remains an open problem, particularly in the intermediate regime where the interaction between the Rashomon set constraint and active learning is most complex.*