We are most grateful to the reviewers for their careful reading of our paper and for their comments.

We agree with the reviewers that the setting depicted in Figure 1 is simplistic. It is definitely not a focus of the paper. We believe our paper stands on its own without it. Of course, if the reviewers have a suggestion for a different figure, it would be most welcome.

Reviewer 3 has a concern about novelty. However, the concern appears to be based on an incorrect premise that the main technical ideas of our paper are published in the prior work of Muthukumar et al. [2020]. On the contrary:

1. The key lemma (Lemma 1) is entirely new, as are the main proof techniques for the main theorems. These new techniques obtain significantly improved conditions for all training points becoming support vectors.
2. The results for the Haar model and the converse result are also entirely new. Muthukumar et al. [2020] has no results for these cases.
3. The only results from Muthukumar et al. [2020] that are quantitatively comparable to ours (1) are restricted to independent and isotropic features, and (2) apply only to the "fixed labels" setting. (See Section 2.2 for descriptions of the data models.)

Reviewer 3 also has a concern about significance due to the focus on hard-margin SVM (as opposed to soft-margin SVM) and the distributional assumptions. Indeed, these are essential ingredients of our result, but we don't see them as major limitations, as we explain below.

1. The hard-margin SVM is of particular interest because it is the limiting solution that arises from gradient-based optimization on the cross-entropy loss on separable training data [Ji and Telgarsky, 2019, Soudry et al., 2018].
2. While the theory of distribution-free learning has been a major achievement in machine learning [Vapnik, 1982, Valiant, 1984], it is not able to explain the generalization behavior of several important machine learning algorithms, including those that produce predictive models that interpolate training data. Indeed, it has long been known that distribution-free learnability is characterized by the VC dimension of the hypothesis class [Vapnik and Chervonenkis, 1971, Blumer et al., 1989], but the VC dimension of linear classifiers in the regime we consider is $\Omega(n \log n)$, where n is the sample size. Hence, in this regime, distribution-free generalization bounds are uninformative.
   The theoretical study of generalization since (at least) the late 1990s has largely shifted to understanding the role of distribution-specific properties (e.g., margins, benign distributions) in order to better understand the behavior of machine learning algorithms. For a variety of examples in the context of SVMs and related algorithms, see [Bartlett and Shawe-Taylor, 1999, Kalai et al., 2008, Talwar, 2020]. Our work "follows in the footsteps" of these prior works in order to study interpolation in very high-dimensional feature spaces.

We are grateful for the opportunity to provide these clarifications, and we hope that the concerns of Reviewer 3 are adequately addressed by these remarks.

Peter Bartlett and John Shawe-Taylor. Generalization performance of support vector machines and other pattern classifiers. In *Advances in Kernel Methods: Support Vector Learning*. MIT Press, Cambridge, MA, 1999.

Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1989.

Ziwei Ji and Matus Telgarsky. The implicit bias of gradient descent on nonseparable data. In *COLT*, 2019.

Adam Tauman Kalai, Adam R Klivans, Yishay Mansour, and Rocco A Servedio. Agnostically learning halfspaces. *SIAM Journal on Computing*, 37(6):1777–1805, 2008.

Vidya Muthukumar, Adhyyan Narang, Vignesh Subramanian, Mikhail Belkin, Daniel Hsu, and Anant Sahai. Classification vs regression in overparameterized regimes: Does the loss function matter? *arXiv:2005.08054*, 2020.

Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Suriya Gunasekar, and Nathan Srebro. The implicit bias of gradient descent on separable data. *Journal of Machine Learning Research*, 19(70):1–57, 2018.

Kunal Talwar. On the error resistance of hinge-loss minimization. In *NeurIPS (to appear)*, 2020.

Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

Vladimir Naumovich Vapnik. *Estimation of dependences based on empirical data*. Springer-Verlag, 1982.

Vladimir Naumovich Vapnik and Alexey Yakovlevich Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971.