Key contributions
The paper presents findings on double descent, where when model size is increased, performance first decreases then increases. They present empirical evidence that challenges and reconciles some of the conventional wisdoms of statistical theories and machine learning. They show that when model complexity is small compared to the number of samples, test error as a function of model complexity behaves like the classical bias/variance tradeoff. However, when model complexity large enough to interpolate, increasing complexity only decreases test error.

Strengths
The paper thoroughly presents experiments and finds that back their hypothesis and findings such as when the model complexity is at the transition from under to over parameterization, increasing the number of samples shifts the peak to the right and actually increases test error. The paper's findings provide useful ways of thinking and explanation for model complexity and performance.

Weaknesses
Many of the paper's findings cease to exist with optimal early stopping. Thus, models with good early stopping will not find much benefit from the findings of the paper. The paper does not provide a full understanding of optimal early stopping and double descent.

My Takeaways

This fact that most of the paper’s findings are negated in the presence of optimal early stopping makes me wonder how to achieve optimal early stopping. With different data and different architectures, optimal early stopping should be related to model complexity in relation to number of samples.