View Reviews

Paper ID
8091

Paper Title
Exact expressions for double descent and implicit regularization via surrogate random design

Reviewer #1

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
This paper studies the problem of double descent for linear regression models. It uses a random design matrix in which the rows are not independent, but are sampled instead from a determinantal point process. This leads to exact expressions for the risk as well as some exact expression for implicit regularization.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
The paper deals with a relevant and timely topic, and it is very well written. The significance is in illustrating how a simple change in the design matrix can lead to exact expressions. The claims are illustrated with empirical evaluations.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
I do not see any big weakness in this work.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?
The methodology seems correct

5. Clarity: Is the paper well written?
Yes, the presentation is very clear and the notation easy to follow.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?
The discussion of previous works is very thorough.

7. Reproducibility: Are there enough details to reproduce the major results of this work?
Yes

9. Please provide an "overall score" for this submission.
7: A good submission; accept.

10. Please provide a "confidence score" for your assessment of this submission.
3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?
Yes

Reviewer #2

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.

This paper proposes closed-form expressions for the mean squared error of least square regression (Moore-Penrose version), when the data comes is sampled from a determinantal point process. It is also proved that this DPP can be a surrogate model for least squares when the data is i.i.d. sampled from a sub-Gaussian distribution (which is more common). Indeed, the difference between the two MSE goes to zero when both the sample size and the dimension go to infinity in a fixed ratio. As a consequence, the paper recovers and improves known results in the double descent literature. In the process, a new mathematical concept of its own interest is introduced: determinant-preserving random matrices.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.

The result is novel.The assumptions are minimal: general position is virtually always satisfied if the distribution has a density, and homoscedastic is not very demanding. I like the idea of surrogate model: using a random model for X for which we can compute more than in the i.i.d. case is appealing.

3. Weaknesses: Explain the limitations of this work along the same axes as above.

I think this is a very good paper, there are no weaknesses so to speak.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?

As far as I can tell, the proofs are correct.

Something that does not seem accurate is the discussion about HMRT19, lines 169-174. Unless I missed something, Figure 2 in the present paper is a plot of the mean squared error MSE=\expec{\norm{\what-\wstar}^2}, as defined line 67. Whereas Figure 2 in HMRT19 is a plot of the generalization risk R=\expec{\norm{X\what-X\wstar}^2}. While these two quantities are related, I fail to see how Figure 2 would contradict the claims in HMRT19.

5. Clarity: Is the paper well written?

The paper is very well written.

typo line 29: thershold
\kappa is not defined in Figure 2

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?

The prior work is discussed extensively.

7. Reproducibility: Are there enough details to reproduce the major results of this work?

Yes

8. Additional feedback, comments, suggestions for improvement and questions for the authors:

Is there any hope that this surrogate design can be used in other situations? Or is this really specific to least squares regression with Moore-Penrose formulation?

9. Please provide an "overall score" for this submission.

9: Top 15% of accepted NeurIPS papers. An excellent submission; a strong accept.

10. Please provide a "confidence score" for your assessment of this submission.

4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?
Yes

Reviewer #3

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
This paper is concerned with the double descent phenomenon. Therefore a linear regression is studied. The goal of the paper is to calculate the mean squared error (MSE) in a non-asymptotic regime.

This is achieved by constructing surrogate random design using the theory of determinantal point processes (DPP). The advantage of this surrogate random design is that the MSE can be calculated exactly. It is shown that the MSE of the original model converges asymptotically with probability one to the MSE of the surrogate random design.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
Recently, the double descent phenomenon has attracted a lot of attention in the ML community. Computing the MSE with high probability non-asymptotically is an important problem. Although this problem is just partially resolved (Conjecture 1) is still open, I think that this paper makes an important step. Also proving the connection with DPPs is interesting.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
-In my opinion, Theorem 3 is the main result in this paper. I think the message of the paper would be much clearer, if Theorem 1 would be put before Theorem 1 and 2. (Theorem 1 is a tool to prove Theorem 3 and its importance becomes only clear through this connection.

-Obviously, the paper would be much stronger if Conjecture 1 would have been proven.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?
The proofs seem to be sound.

5. Clarity: Is the paper well written?
The clarity of writing could be improved:

-The main theorems could be stated more clearly. In particular, the authors could state the assumptions more precisely. For example, in Theorem 1 it could be mentioned how w is chosen. (I guess the authors mean an arbitrary w.) Since people do not read papers from beginning to end, this would increase the readability a lot.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?
Prior work is appropriately described.

7. Reproducibility: Are there enough details to reproduce the major results of this work?
Yes

8. Additional feedback, comments, suggestions for improvement and questions for the authors:
-typo l. 29: "interpolation threshold"

l. 198: "product of non-zero eigenvalues". It would be great if the authors could clarify how eigenvalues with algebraic multiplicity larger than one are counted.

-Broader impact statement: "Our research can be applied here to provide a theoretical understanding of the surprising phenomenon where more data leads to worse generalization performance." I do not see where this statement is supported in the paper. It also seems to contradict the (recent) paper [1]. It would be great if the authors either explain it better or remove it from the broader impact statement.

[1]: arxiv.org/abs/2003.01897 Optimal Regularization Can Mitigate Double Descent
Preetum Nakkiran, Prayaag Venkat, Sham Kakade, Tengyu Ma

9. Please provide an "overall score" for this submission.

7: A good submission; accept.

10. Please provide a "confidence score" for your assessment of this submission.

3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?

Yes