UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTE OF COMPUTING

SCIENTIFIC METHODOLOGY FOR COMPUTING
MO430

EXERCISE 1

STUDENT: Gian Franco Joel Condori Luna
RA: 234826

SEARCH BIBLIOGRAPHIC

1.- Look for bibliography on "double descent" which is a recent phenomenon in machine learning, and has to do with "deeplearning", "bias" "validation" "training".
2.- List 7 articles that you think are the most important for you to read to learn about the subject.
3.- Indicate who you think is the main researcher to follow on this subject.

DEVELOPMENT OF EXERCISE

The Double Descent Hypothesis: How Bigger Models and More Data Can Hurt Performance

Bigger is better certainly applies to modern deep learning paradigms. Large neural networks with millions of parameters have regularly outperformed collections of smaller networks specialized on a given task. Some of the most famous models of the last few years such as Google BERT, Microsoft T-NLG or OpenAI GPT-2 are so large that their computational cost results prohibited for most organizations. However, the performance of a model does not increase linearly with its size. Double descent is a phenomenon where, as we increase model size, performance first gets worse and then gets better. Recently, OpenAI researchers studied how many modern deep learning models are vulnerable to the double-descent phenomenon.

The relationship between the performance of a model and its size have certainly puzzled deep learning researchers for years. In traditional statistical learning, the bias-variance trade off states that models of higher complexity have lower bias but higher variance. According to this theory, once model complexity passes a certain threshold, models “overfit” with the variance term dominating the test error, and hence from this point onward, increasing model complexity will only decrease performance. From that perspective, statistical learning tells us that *“larger models are worse”*. However, modern deep learning model have challenged this conventional wisdom.

Contrasting with the bias-variance tradeoff, deep neural networks with millions of parameters have proven to outperform smaller models. Additionally, many of these models improve linearly with more training data. Therefore, the conventional wisdom among deep learning practitioners is that *“larger models and more data are always better”*.

Which theory is correct? Statistical learning or the empirical evidence of deep learning models? The double descent phenomenon show us that both theories can be reconciled but, also, that some of its assumptions are erroneous.

List of articles:

1.-

- **Title:** Deep Double Descent: Where bigger models and more data hurt.
- **Authors:**
    - Preetum Nakkiran(Harvard University).
    - Girl Kaplun(Harvard University).
    - Yamini Bansal(Harvard University).
    - Tristan Yang(Harvard University).
    - Boaz Barak(Harvard University).
    - Ilya Sutskever(OpenAI).
- **Link:** https://arxiv.org/pdf/1912.02292.pdf

2.-

- **Title:** On the Role of Optimization in Double Descent: A Least Squares Study
- **Authors:**
    - Ilja Zbarszkij (DeepMind).
    - Csaba Szepesvári (DeepMind, Canada
    - University of Alberta, Edmonton).
    - Omar Rivasplata (University College London).
    - Amal Rannen-Triki (DeepMind).
    - Razvan Pascanu (DeepMind).
- **Link:** https://arxiv.org/pdf/2107.12685.pdf

3.-

- **Title:** Mitigating deep double descent by concatenating inputs
- **Authors:**
    - John Chen.
    - Qihan Wang.
    - Anastasios Kyrillidis.
- **Link:** https://arxiv.org/pdf/2107.00797.pdf

4.-

- **Title:** Double Descent and Other Interpolation Phenomena in GANs.
- **Authors:**
    - Lorenzo Luzi (Rice University).
    - Yehuda Dar (Rice University).
    - Richard G. Baraniuk (Rice University).
- **Link:** https://arxiv.org/pdf/2106.04003.pdf

5.-

- **Title:** Double Descent Optimization Pattern and Aliasing: Caveats of Noisy Labels.
- **Authors:**
    - Florian Dubost (Stanford University).

    - Khaled K. Saab (Stanford University).
    - Erin Hong (Stanford University).
    - Daniel Y. Fu (Stanford University).
    - Max Pike (Stanford University).
    - Siddharth Sharma (Stanford University).
    - Siyi Tang (Stanford University).
    - Nandita Bhaskhar (Stanford University).
    - Christopher Lee-Messer (Stanford University).
    - Daniel Rubin (Stanford University).
- **Link:** https://arxiv.org/pdf/2106.02100.pdf

6.-

- **Title:** Transfer Learning Can Outperform the True Prior in Double Descent Regularization
- **Authors:**
    - Yehuda Dar.
    - Richard G. Baraniuk
- **Link:** https://arxiv.org/pdf/2103.05621.pdf

7.-

- **Title:** A brief prehistory of double descent.
- **Authors:**
    - Marco Loog.
    - Tom Viering.
    - Alexander Mey.
    - Jesse H. Krijthe.
    - David M. J. Tax
- **Link:** https://www.pnas.org/content/117/20/10625