Machine learning in production

References

Introduction to Machine Learning in Production

This is a compilation of resources including URLs and papers appearing in lecture videos.

Overall resources:

Konstantinos, Katsiapis, Karmarkar, A., Altay, A., Zaks, A., Polyzotis, N., … Li, Z. (2020). Towards ML Engineering: A brief history of TensorFlow Extended (TFX). <http://arxiv.org/abs/2010.02013>

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2020). Challenges in deploying machine learning: A survey of case studies. <http://arxiv.org/abs/2011.09926>

Week 1: Overview of the ML Lifecycle and Deployment

[Concept and Data Drift](https://towardsdatascience.com/machine-learning-in-production-why-you-should-care-about-data-and-concept-drift-d96d0bc907fb)

[Monitoring ML Models](https://christophergs.com/machine%20learning/2020/03/14/how-to-monitor-machine-learning-models/)

[A Chat with Andrew on MLOps: From Model-centric to Data-centric AI](https://youtu.be/06-AZXmwHjo)

Konstantinos, Katsiapis, Karmarkar, A., Altay, A., Zaks, A., Polyzotis, N., … Li, Z. (2020). Towards ML Engineering: A brief history of TensorFlow Extended (TFX). <http://arxiv.org/abs/2010.02013>

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2020). Challenges in deploying machine learning: A survey of case studies. <http://arxiv.org/abs/2011.09926>

Sculley, D., Holt, G., Golovin, D., Davydov, E., & Phillips, T. (n.d.). Hidden technical debt in machine learning systems. Retrieved April 28, 2021, from Nips.cc

<https://papers.nips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf>

Week 2: Select and Train Model

[Establishing a baseline](https://blog.ml.cmu.edu/2020/08/31/3-baselines/)

[Error analysis](https://techcommunity.microsoft.com/t5/azure-ai/responsible-machine-learning-with-error-analysis/ba-p/2141774)

[Experiment tracking](https://neptune.ai/blog/ml-experiment-tracking)

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., … Anderljung, M. (n.d.). Toward trustworthy AI development: Mechanisms for supporting verifiable claims∗. Retrieved May 7, 2021 [http://arxiv.org/abs/2004.07213v2](http://arxiv.org/abs/2004.07213v2" \o "" \t "_blank)

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Sutskever, I. (2019). Deep double descent: Where bigger models and more data hurt. Retrieved from <http://arxiv.org/abs/1912.02292>

Week 3: Data Definition and Baseline

[Label ambiguity](https://csgaobb.github.io/Projects/DLDL.html)

<https://arxiv.org/pdf/1706.06969.pdf>

[Data pipelines](https://cs230.stanford.edu/blog/datapipeline/)

[Data lineage](https://blog.tensorflow.org/2021/01/ml-metadata-version-control-for-ml.html)

[MLops](https://cloud.google.com/blog/products/ai-machine-learning/key-requirements-for-an-mlops-foundation)

Geirhos, R., Janssen, D. H. J., Schutt, H. H., Rauber, J., Bethge, M., & Wichmann, F. A. (n.d.). Comparing deep neural networks against humans: object recognition when the signal gets weaker∗. Retrieved May 7, 2021, from Arxiv.org website: