ENGAGE: NExt GeNeration ComputinG Environments for Artificial intelliGEnce

Work Package 2:

Reproducible deployment and scheduling strategies for AI workloads

ML Model Performance Monitoring with Generative Models

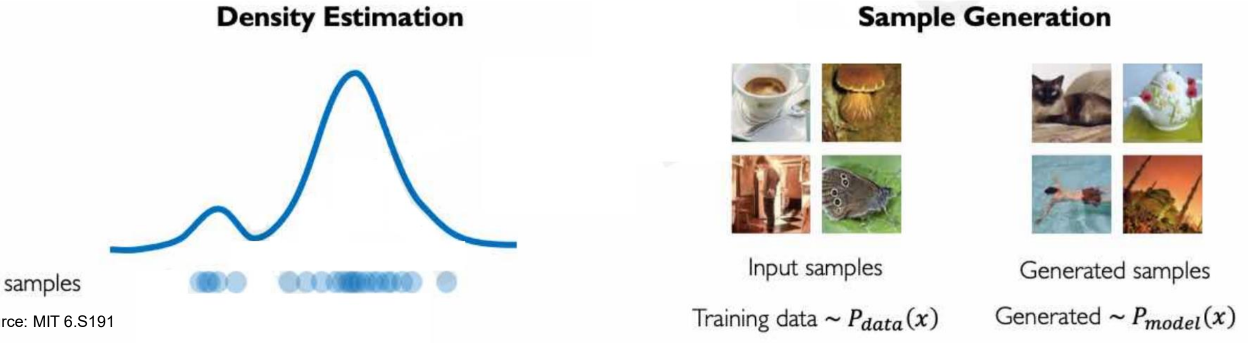
Melvin Chelli

Objectives:

- **Detect** model performance degradation after deployment.
- Investigate **causes** for this
- Formulate **automated retraining** strategies to compensate

Generative Models

- Take some training samples from some distribution as input and learn a model that represents that distribution.
- If training data P_{data} (x), and we have a model that is from P_{model}(x), the goal is to find P_{model}(x) similar to P_{data}(x)
- Simple example: text prediction models.



Density Estimation

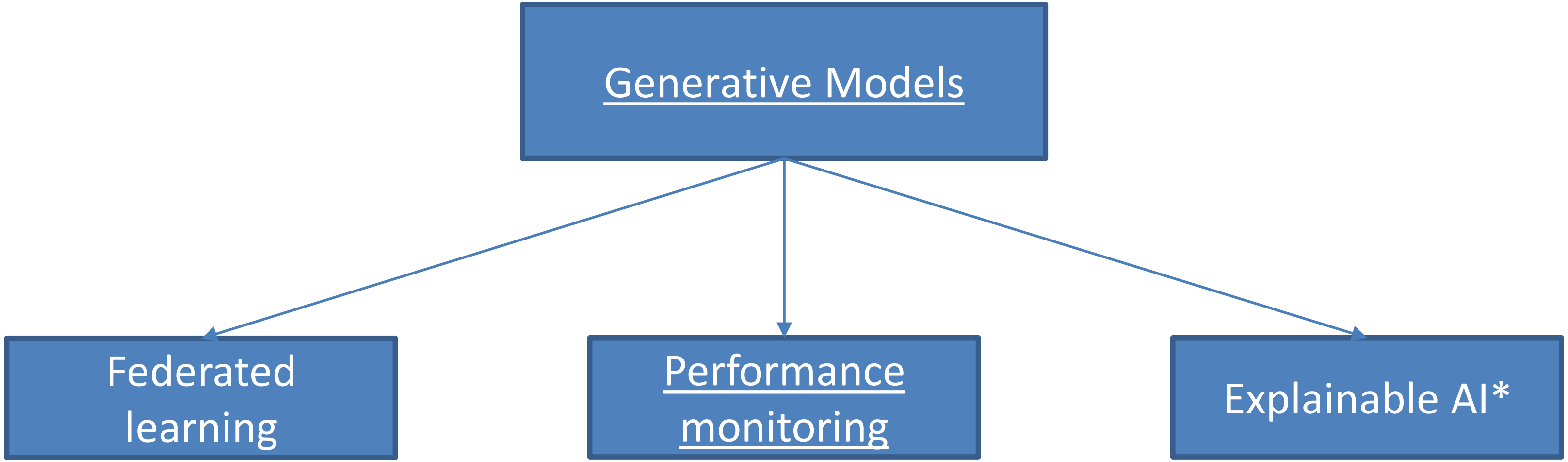
Sample Generation

samples
Input samples
Generated samples
Training data $\sim P_{data}(x)$
Generated $\sim P_{model}(x)$

Images source: MIT 6.S191

Motivation:

- Why generative models?



Generative Models

Federated learning

Performance monitoring

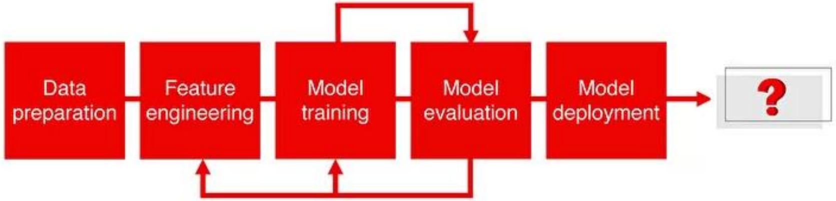
Explainable AI*

*Angelov, Plamen, and Eduardo Soares. "Towards explainable deep neural networks (xDNN)." *Neural Networks 130* (2020): 185-194.

Motivation: Performance Monitoring

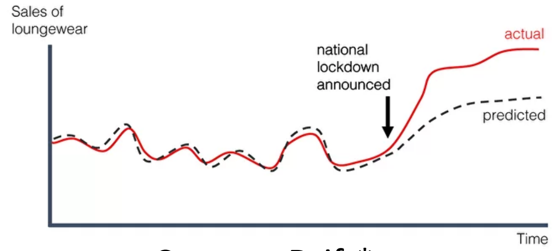
German Research Center for Artificial Intelligence

- What is performance?
- Deep neural networks are often trained with **closed-world assumption** i.e the test data distribution is assumed to be similar to the training data distribution.
- However, when employed in real-world tasks, this assumption doesn't hold true leading to a significant drop in their performance.
- An ideal AI system flag the ones that are beyond its capability to seek human intervention.
- Monitoring!

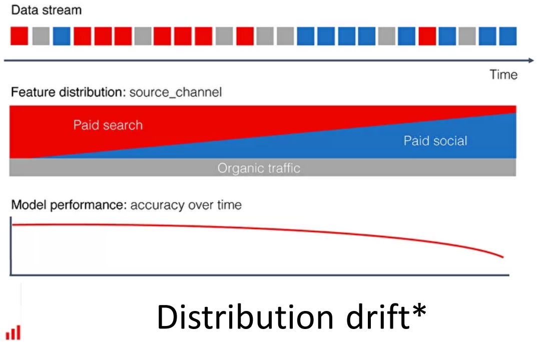


Performance monitoring

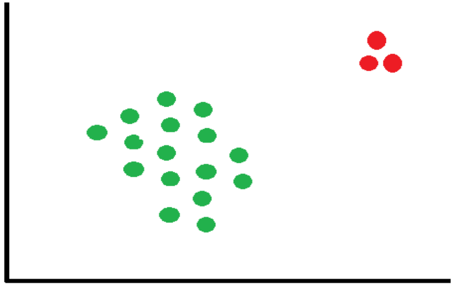
- Causes for performance decay:
  - Concept and distribution drift
  - Outliers
  - Adversarial attacks



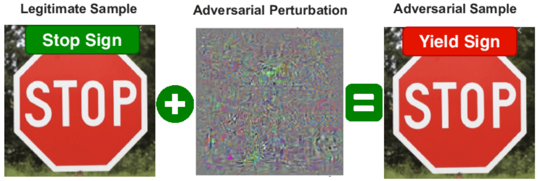
Concept Drift*



Distribution drift*



Outliers



Adversarial attacks[†]

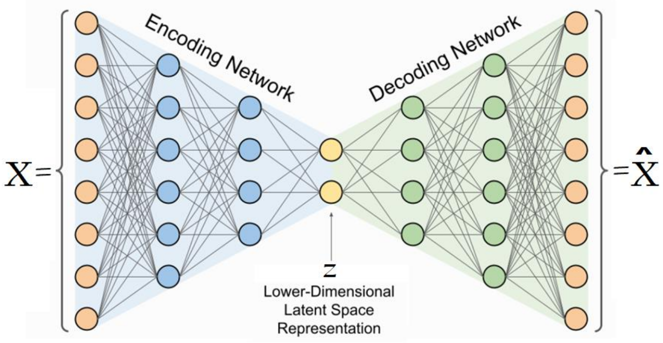
Sources:

* Machine Learning Monitoring, Part 5, www.evidentlyai.com,

[†] Ahmad et al. (2020), Developing Future Human-Centered Smart Cities: Critical Analysis of Smart City Security, Interpretability, and Ethical Challenges

Work Done:

- As a first step, closer look at reconstruction-based methods [1,2].
- Autoencoders:
- Loss: $\mathcal{L}(x, \hat{x}) = \| x - \hat{x} \|^2$



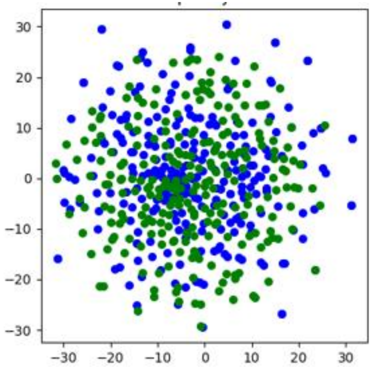
Source: www.assemblyai.com

[1] Yang, J., Zhou, K., Li, Y., & Liu, Z. 2021. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*.

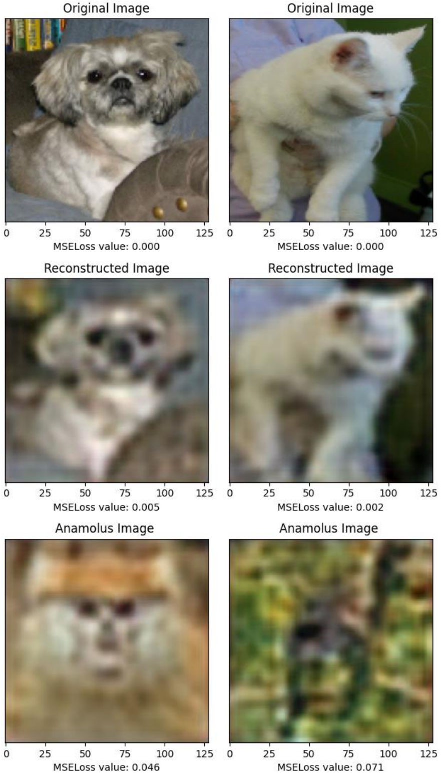
[2] Kieu, T., Yang, B., Guo, C., Jensen, C. S., Zhao, Y., Huang, F., & Zheng, K. (2022). Robust and explainable autoencoders for time series outlier detection. In Proceeding of the 38th IEEE International Conference on Data Engineering, ICDE 2022.

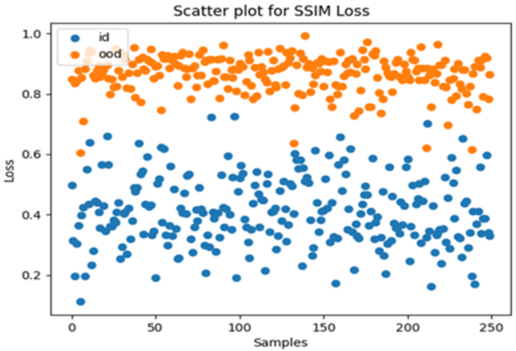
Work Done:

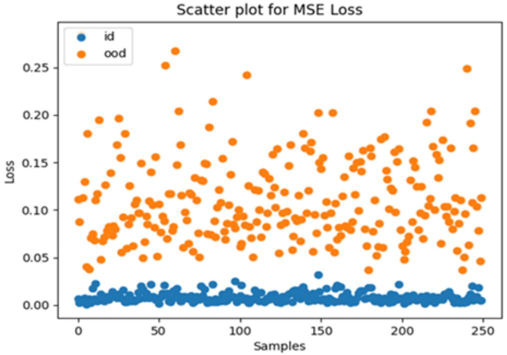
- Some first implementations and validation of autoencoders as outlier detecting tools
- Good performance for outlier detection



Latent space visualized

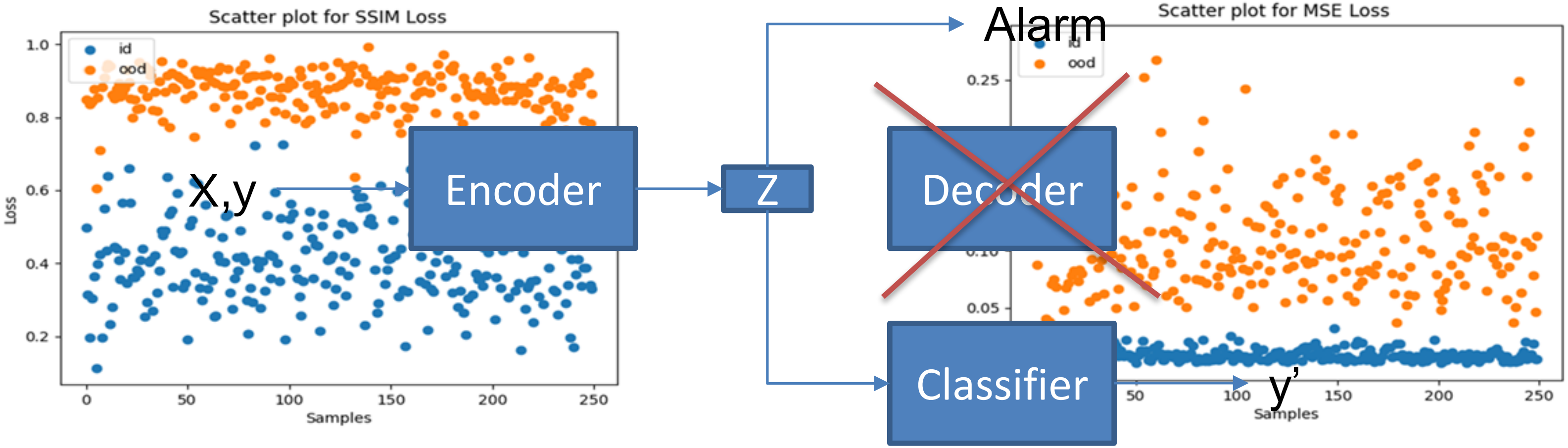






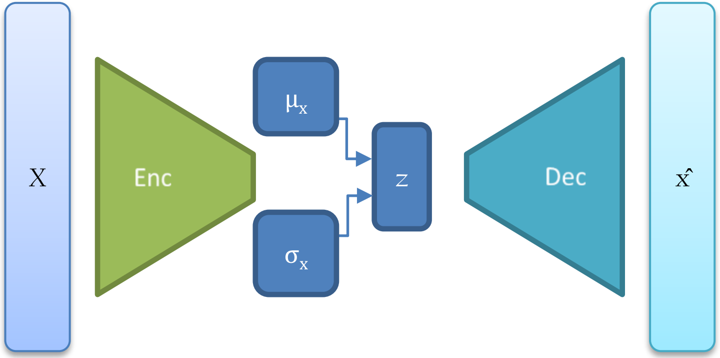
Interesting questions:

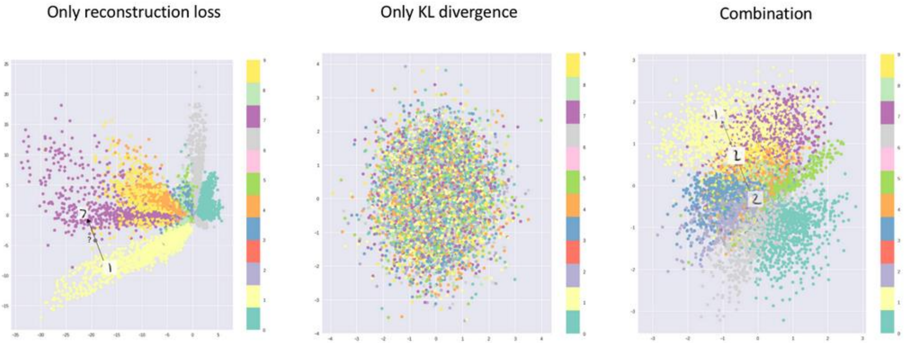
- What are the best loss metrics?
- Can the latent space provide sufficient information to be leveraged for outlier detection?
- How about a training strategy that trains autoencoder (for anomaly detection) and classifier together, and later dispose the decoder?



Variational autoencoder

- Vanilla autoencoders cannot produce a latent space that is regularized. Outlier detection using latent space is not possible.
- Variational autoencoders attempt to solve this.
- Work in progress.



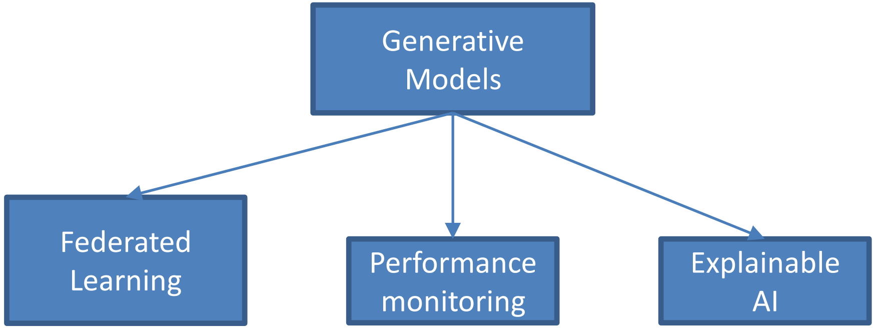


Source: **Intuitively Understanding Variational Autoencoders**
www.towardsdatascience.com

- Loss: $\mathcal{L}(\phi, \theta, x) = (reconstruction\ loss) + (regularization\ term)$

Beyond monitoring?

- Federated learning conventionally uses models with fixed architecture, and only adapts weights and biases. The idea is that the data seen by local models cannot be shared, i.e the data on which model is improved is distributed.
- If something drastic occurs at the model, in order to change the underlying model architecture, additional central training data is needed.
- Here, generative models can be used to generate plausible realistic data, train the networks, and propagate the new architecture.


Generative Models

Federated Learning

Performance monitoring

Explainable AI

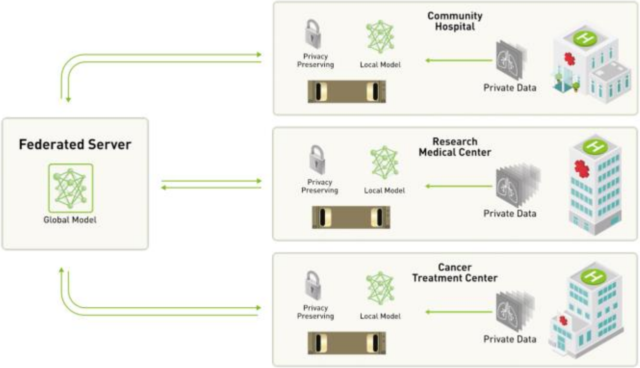


Image source: https://blogs.nvidia.com/blog/2019/10/13/what-is-federated-learning/

References

[1] Yang, J., Zhou, K., Li, Y., & Liu, Z. 2021. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*.

[2] Kieu, T., Yang, B., Guo, C., Jensen, C. S., Zhao, Y., Huang, F., & Zheng, K. (2022). Robust and explainable autoencoders for time series outlier detection. In Proceeding of the 38th IEEE International Conference on Data Engineering, ICDE 2022.

[3] Bulusu, S., Kailkhura, B., Li, B., Varshney, P.K., Song, D. (2020). Out-of-Distribution Detection in Deep Learning: A Survey. *Preprint.*

[4] Schelter, S., Rukat, T., and Biessmann, F. 2020. Learning to Validate the Predictions of Black Box Classifiers on Unseen Data. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD '20)*.

[5] Lipton, Z., Wang, Y. X., & Smola, A. 2018. Detecting and correcting for label shift with black box predictors. In *International conference on machine learning*.

[6] Krawczyk, B., Pfahringer, B., Wozniak, M. (2018). Combining active learning with concept drift detection for data stream mining. 2239-2244. 10.1109/BigData.2018.8622549.

References

[7] Jaworski, M., Rutkowski, L., Angelov, P. (2020). Concept Drift Detection Using Autoencoders in Data Streams Processing. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds) Artificial Intelligence and Soft Computing. ICAISC 2020. Lecture Notes in Computer Science, vol 12415. Springer, Cham.

[8] Cerqueira, V., Gomes, H.M., Bifet, A. (2020). Unsupervised Concept Drift Detection Using a Student–Teacher Approach. In: Appice, A., Tsoumakas, G., Manolopoulos, Y., Matwin, S. (eds) Discovery Science. DS 2020. Lecture Notes in Computer Science, vol 12323. Springer, Cham. https://doi.org/10.1007/978-3-030-61527-7_13

[9]Ahmad, Kashif & Maabreh, Majdi & Ghaly, Mohamed & Khan, Khalil & Qadir, Junaid & Al-Fuqaha, Ala. (2020.). Developing Future Human-Centered Smart Cities: Critical Analysis of Smart City Security, Interpretability, and Ethical Challenges

[10] Angelov, Plamen, and Eduardo Soares. "Towards explainable deep neural networks (xDNN)." *Neural Networks 130* (2020): 185-194.