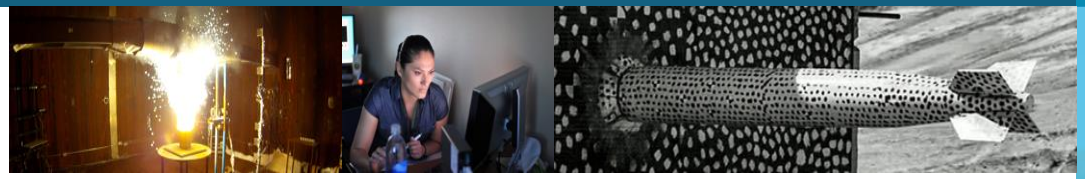
SAND2021-12358PE

Sandia National Laboratories

Likelihood Ratios for Out-of-Distribution Detection

Ren J., Liu P., Fertig, E., Snoek, J., Poplin, R., DePristo, M., ... Lakshminarayanan, B.

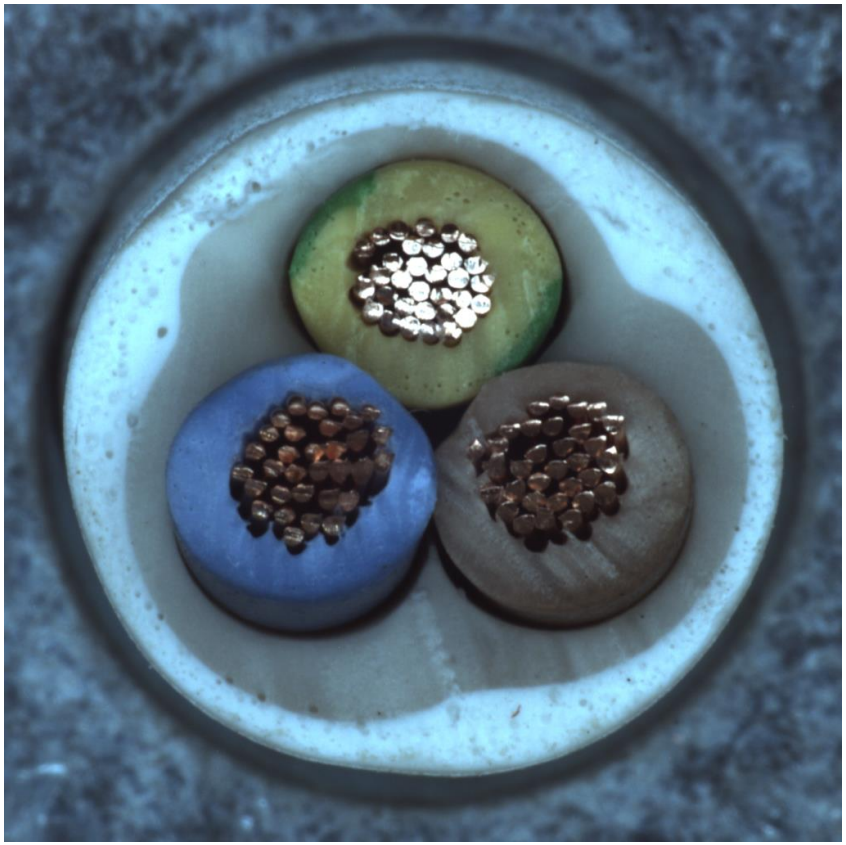


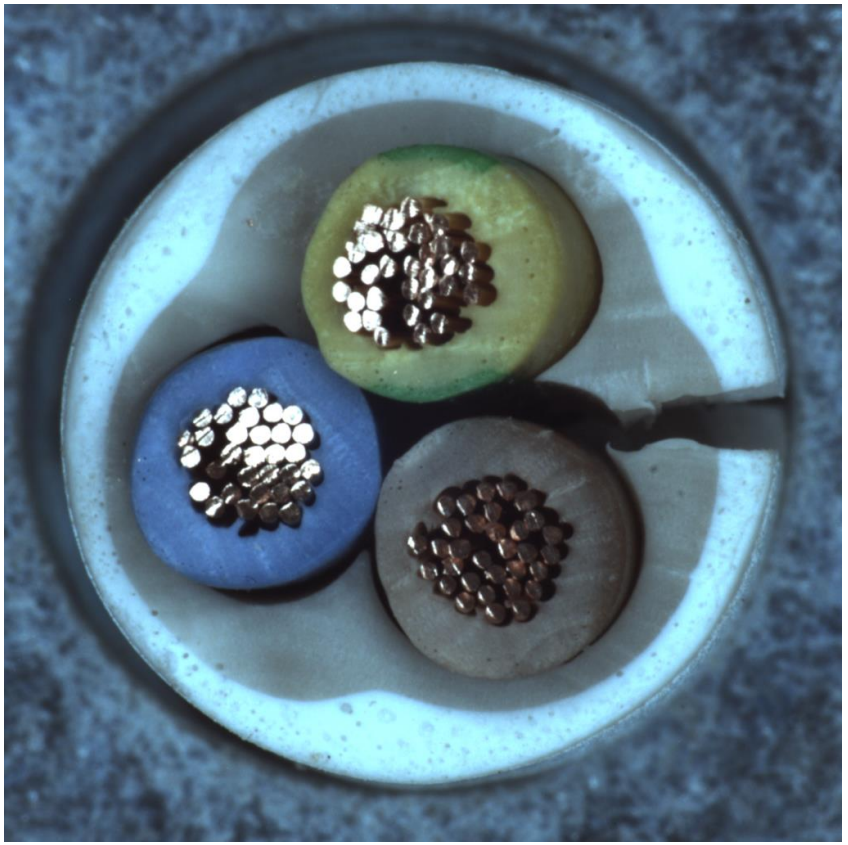
Presented by: Abigail Pribisova

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

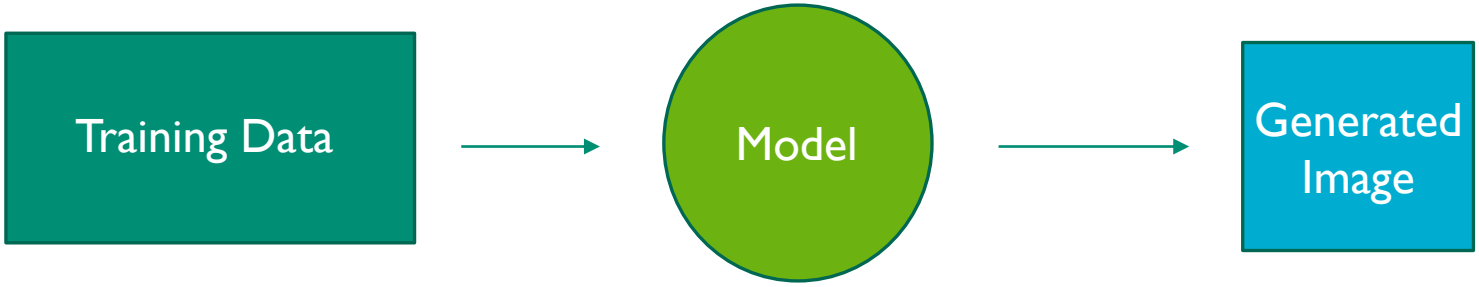
Anomaly Detection





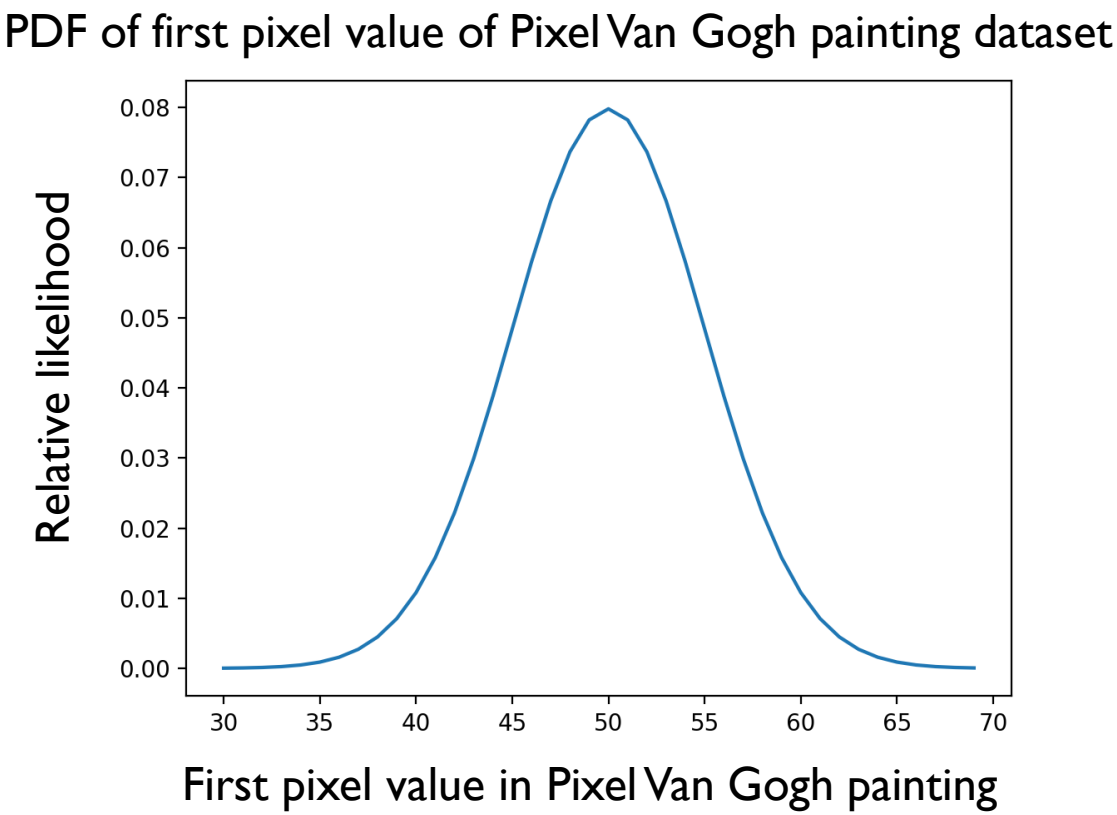
https://www.mvtec.com/company/research/datasets/mvtec-ad/

Generative Models



https://www.linkedin.com/pulse/gans-generative-adversarial-networks-hesham-eraqi

Density Estimation Problem



https://machinelearningmastery.com/continuous-likelihood-distributions-for-machine-learning/
https://www.connox.com/categories/home-decor/graphics-posters/ixxi-van-gogh-pixel.html

Joint PDF

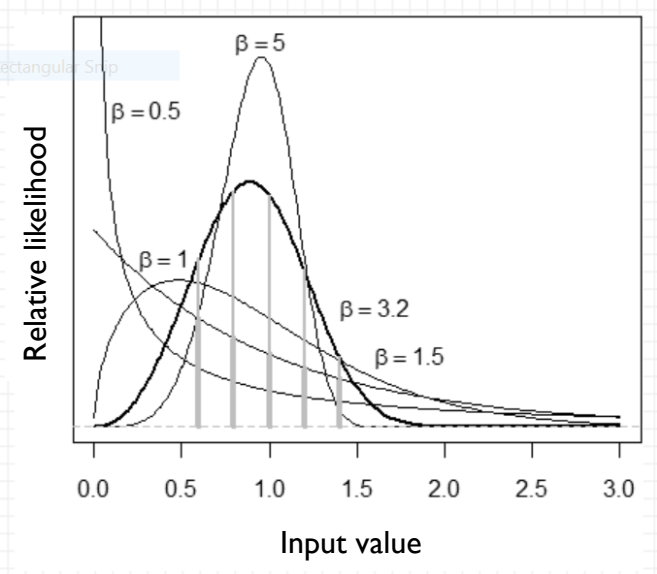
https://www.theurbanlist.com/a-list/best-adult-jigsaw-puzzle

Autoregressive Model

Training Data → Model →

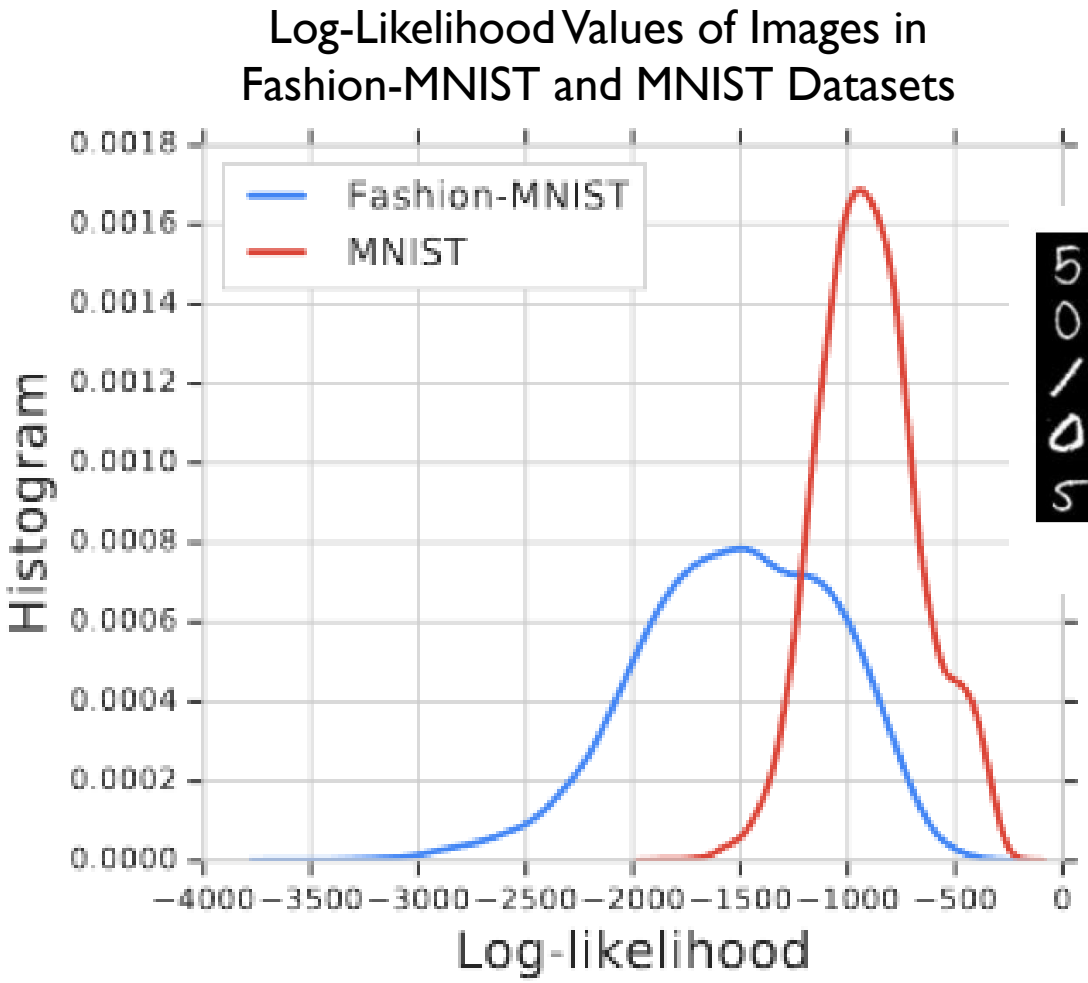
During training:
Is relative likelihood of the training dataset **high** for these parameters?

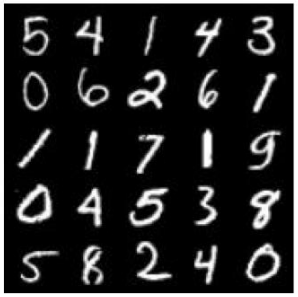
PDF of input distribution with various β parameters



http://www.statisticalengineering.com/Weibull/likelihood.html

Problem





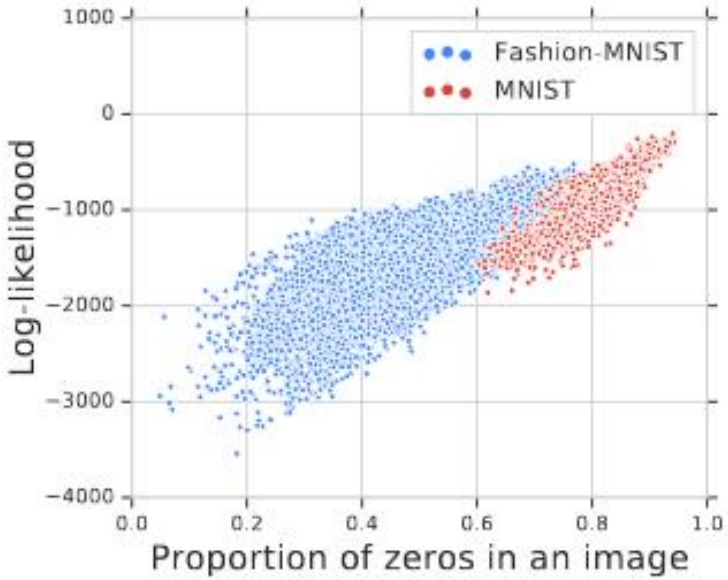
MNIST

Fashion MNIST

https://arxiv.org/pdf/1906.02845.pdf

Fashion MNIST vs. MNIST

Log-Likelihood Values of Images in Fashion-MNIST and MNIST Datasets in Proportion to Zeros in the Images



MNIST

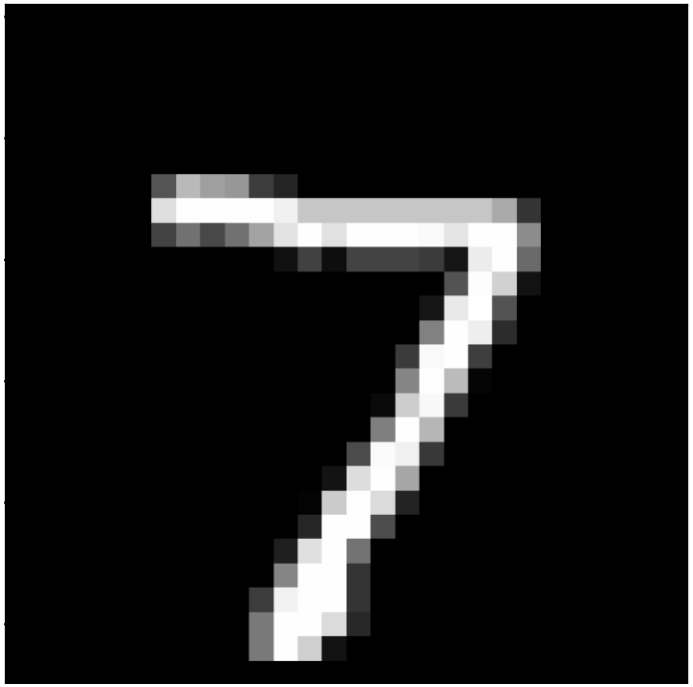
Fashion MNIST

https://arxiv.org/pdf/1906.02845.pdf

https://emiliendupont.github.io/2018/03/14/mnist-chicken/

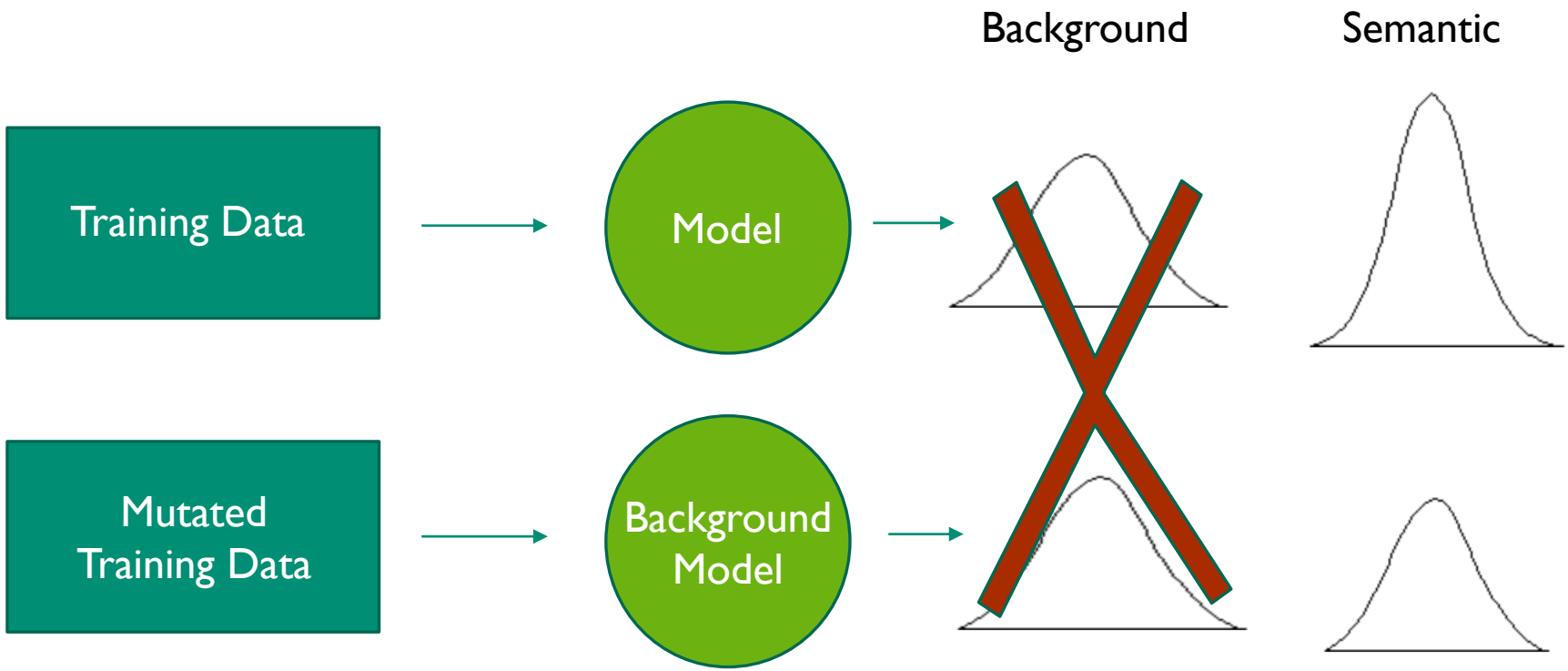
Background vs. Semantic Likelihood

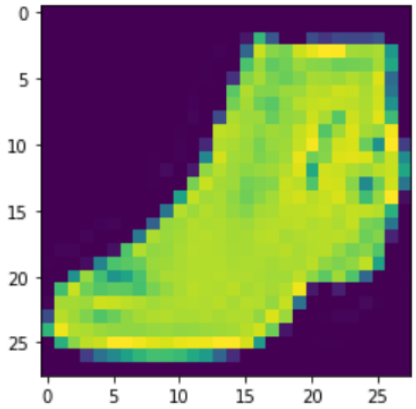
$$p(\boldsymbol{x}) = p(\boldsymbol{x}_B)p(\boldsymbol{x}_S).$$

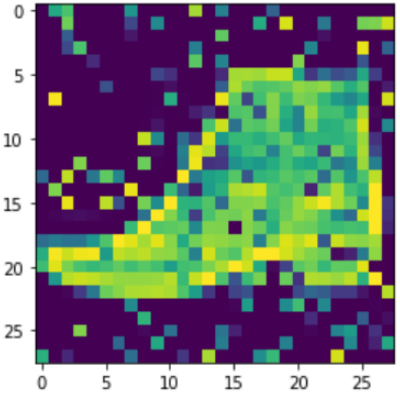


https://arxiv.org/pdf/1906.02845.pdf
https://machinelearningmastery.com/how-to-develop-a-convolutional-neural-network-from-scratch-for-mnist-handwritten-digit-classification/

Proposed Method

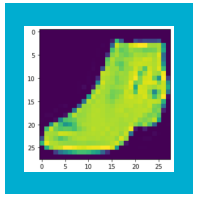


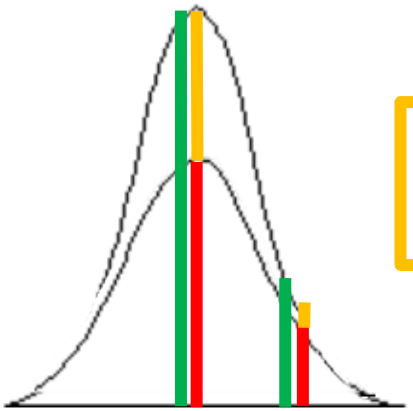




Log-Likelihood Ratio

$$\mathrm{LLR}(\boldsymbol{x}) = \log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_B)\, p_{\boldsymbol{\theta}}(\boldsymbol{x}_S)}{p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_B)\, p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_S)},$$



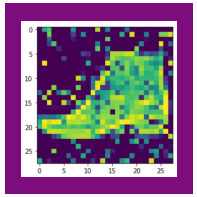


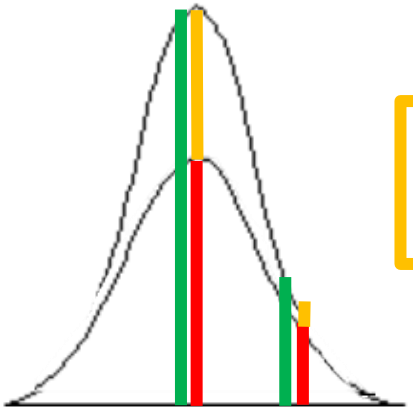
$$\mathrm{LLR}(\boldsymbol{x}) \approx \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_S) - \log p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_S).$$

https://arxiv.org/pdf/1906.02845.pdf
https://docs.oracle.com/cd/E57185_01/CBREG/ch03s02s03s02.html

Log-Likelihood Ratio

$$\mathrm{LLR}(\boldsymbol{x}) = \log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_B)\, p_{\boldsymbol{\theta}}(\boldsymbol{x}_S)}{p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_B)\, p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_S)},$$



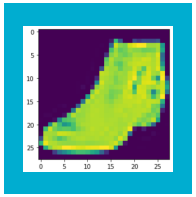


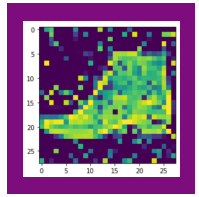
$$\mathrm{LLR}(\boldsymbol{x}) \approx \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_S) - \log p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_S).$$

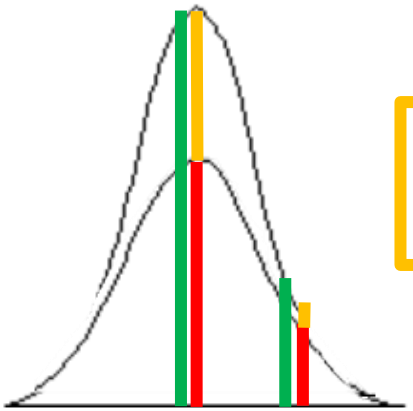
https://arxiv.org/pdf/1906.02845.pdf
https://docs.oracle.com/cd/E57185_01/CBREG/ch03s02s03s02.html

Log-Likelihood Ratio

$$\mathrm{LLR}(\boldsymbol{x}) = \log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_B)\, p_{\boldsymbol{\theta}}(\boldsymbol{x}_S)}{p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_B)\, p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_S)},$$



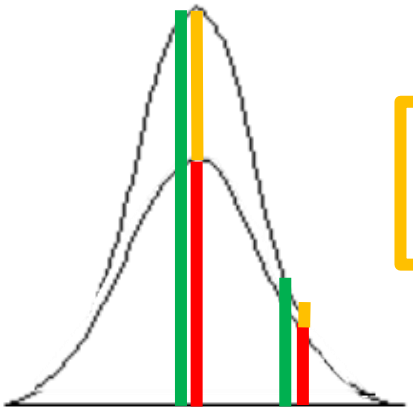




$$\mathrm{LLR}(\boldsymbol{x}) \approx \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_S) - \log p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_S).$$

https://arxiv.org/pdf/1906.02845.pdf
https://docs.oracle.com/cd/E57185_01/CBREG/ch03s02s03s02.html

Log-Likelihood Ratio



$$\mathrm{LLR}(\boldsymbol{x}) \approx \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_S) - \log p_{\boldsymbol{\theta}_0}(\boldsymbol{x}_S).$$

https://arxiv.org/pdf/1906.02845.pdf
https://docs.oracle.com/cd/E57185_01/CBREG/ch03s02s03s02.html

Train & Validate

Train:

Input:

In-distribution dataset

- In-distribution inputs (Fashion MNIST)

Validate:

- In-distribution inputs (Fashion MNIST)
- Out-of-distribution inputs (NotMNIST)

http://yaroslavvb.blogspot.com/2011/09/notmnist-dataset.html

Test & Evaluate

Test:

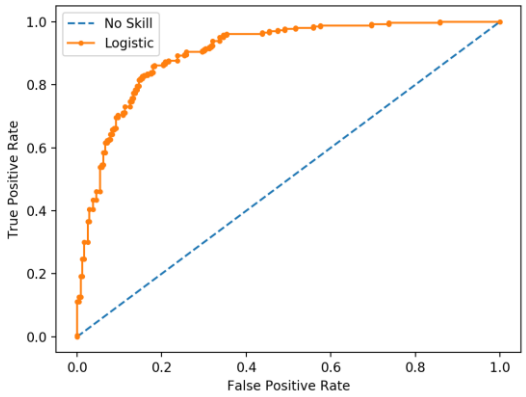
- In-distribution inputs (Fashion MNIST)
- Out-of-distribution inputs (MNIST)

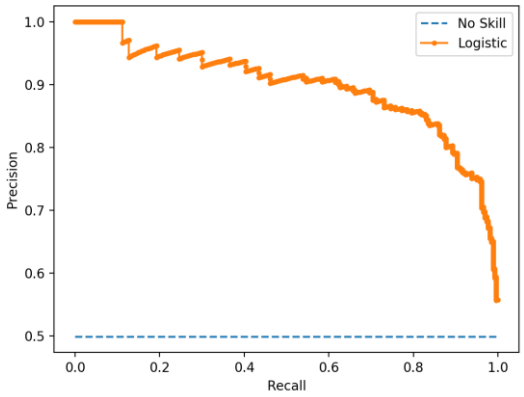
Output:
Log-likelihood ratio for each input

Evaluate:

- Low LLR → Out-of-distribution

Metrics:
AUROC, AUPRC, FPR80

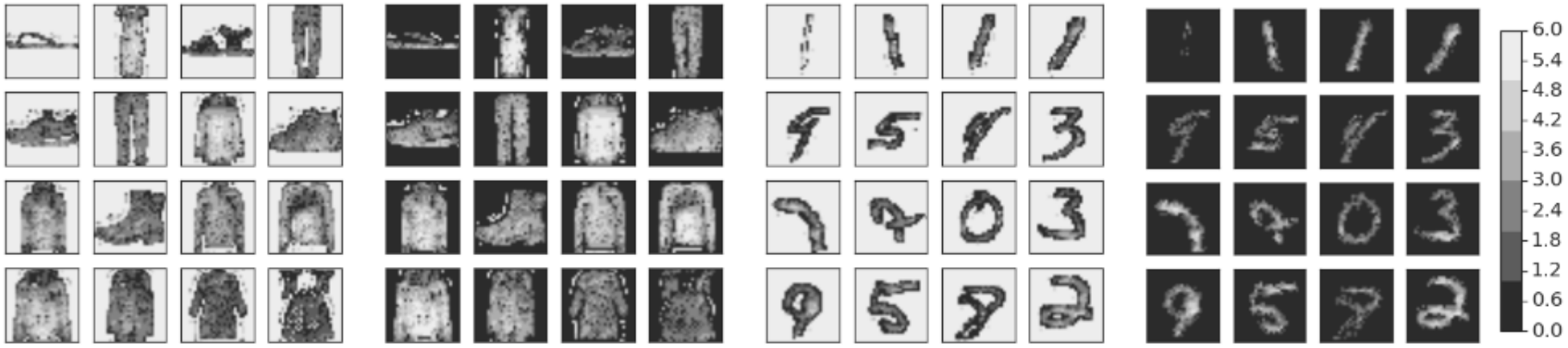




Number of false positives
/
(Number of false positives + number of true negatives)
= FPR

https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-imbalanced-classification

Pixel Heatmap

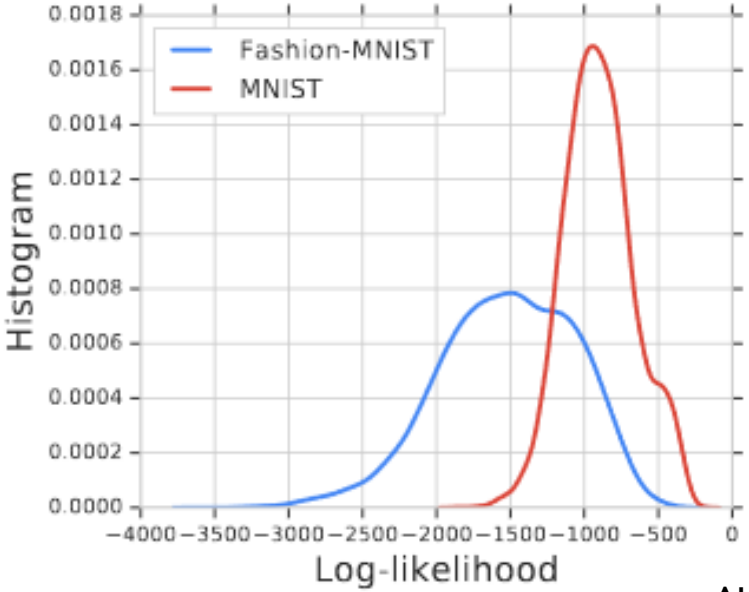


(a) Likelihood (b) Likelihood-Ratio (c) Likelihood (d) Likelihood-Ratio

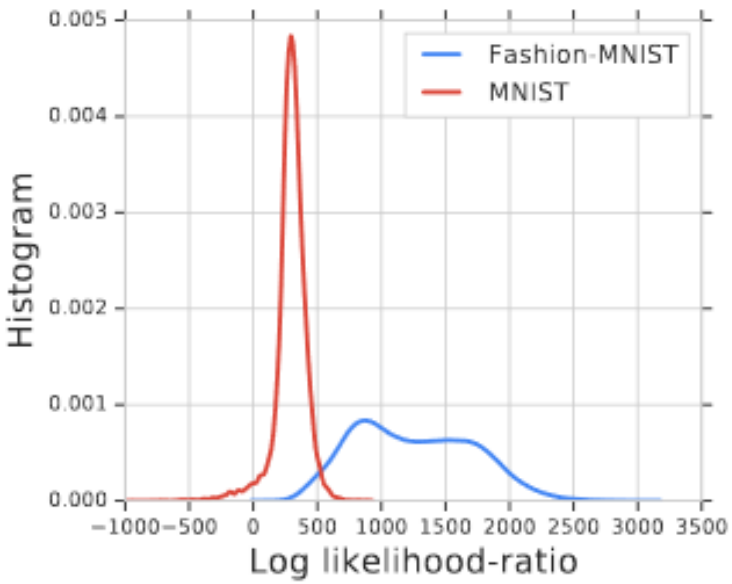
https://arxiv.org/pdf/1906.02845.pdf

Results

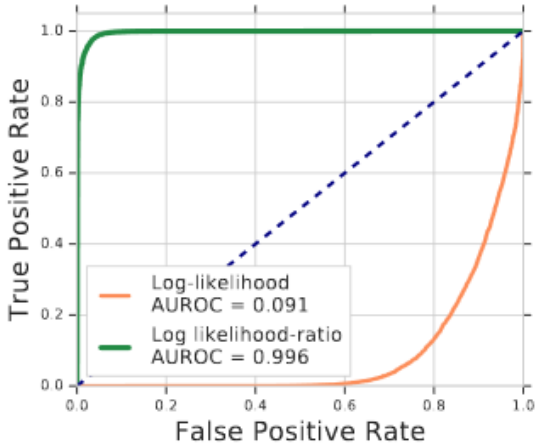
Log-Likelihood Values of Images in Fashion-MNIST and MNIST Datasets



Log-Likelihood Ratio Values of Images in Fashion-MNIST and MNIST Datasets



AUROC for OOD classification



https://arxiv.org/pdf/1906.02845.pdf

Conclusion

- Problem
- Solution
- Result
- Application

https://responsewebrecruitment.co.uk/online-recruitment-blog/100-most-ridiculous-job-interview-questions-ever/
https://twitter.com/sandialabs

Works Cited

Pinaya, W. H. (2020, March 18). Autoregressive Models - PixelCNN. Retrieved from https://towardsdatascience.com/autoregressive-models-pixelcnn-e30734ede0c1

Ren, J. et al. (2019). Likelihood Ratios for Out-of-Distribution Detection. Retrieved July 13, 2020 from https://arxiv.org/pdf/1906.02845.pdf.