به نام ایزد یکتا

دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر

تمرین هفتم درس روش پژوهش و ارائه

استاد: دکتر رضا صفابخش

تهیه کننده: بردیا اردکانیان

برسی و مهار مثال‌های خصمانه در یادگیری ماشین

Bibliography

Nature

1. Tabassi, E., Burns, K. J., Hadjimichael, M., Molina-Markham, A. D. & Sexton, J. T. *A taxonomy and terminology of adversarial machine learning*. https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf (2019) doi:10.6028/NIST.IR.8269-draft.

2. Adversarial Examples Are Not Easily Detected | Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security. https://dl.acm.org/doi/abs/10.1145/3128572.3140444.

3. Wiyatno, R. R., Xu, A., Dia, O. & de Berker, A. Adversarial Examples in Modern Machine Learning: A Review. (2019).

4. Yuan, X., He, P., Zhu, Q. & Li, X. Adversarial Examples: Attacks and Defenses for Deep Learning. *IEEE Transactions on Neural Networks and Learning Systems* **30**, 2805–2824 (2019).

5. Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I. P. & Tygar, J. D. Adversarial machine learning. in *Proceedings of the 4th ACM workshop on Security and artificial intelligence* 43–58 (Association for Computing Machinery, 2011). doi:10.1145/2046684.2046692.

6. Adversarial Machine Learning Attacks and Defense Methods in the Cyber Security Domain | ACM Computing Surveys. https://dl.acm.org/doi/10.1145/3453158.

7. Adversarial machine learning for cybersecurity and computer vision: : Current developments and challenges: WIREs Computational Statistics: Vol 12, No 5. *Wiley Interdisciplinary Reviews: Computational Statistics*.

8. Zheng, H., Zhang, Z., Gu, J., Lee, H. & Prakash, A. Efficient Adversarial Training With Transferable Adversarial Examples. in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 1178–1187 (IEEE, 2020). doi:10.1109/CVPR42600.2020.00126.

9. Goodfellow, I. J., Shlens, J. & Szegedy, C. Explaining and Harnessing Adversarial Examples. (2015).

10. Zhao, Z., Dua, D. & Singh, S. Generating Natural Adversarial Examples. (2018).

11. Zizzo, G., Hankin, C., Maffeis, S. & Jones, K. INVITED: Adversarial Machine Learning Beyond the Image Domain. in *2019 56th ACM/IEEE Design Automation Conference (DAC)* 1–4 (2019).

12. Laskov, P. & Lippmann, R. Machine learning in adversarial environments. *Mach Learn* **81**, 115–119 (2010).

13. McDaniel, P., Papernot, N. & Celik, Z. B. Machine Learning in Adversarial Settings. *IEEE Security Privacy* **14**, 68–72 (2016).

14. Machine learning uncertainties with adversarial neural networks | SpringerLink. https://link.springer.com/article/10.1140/epjc/s10052-018-6511-8.

15. Brachman, R. J. & Stone, P. Synthesis Lectures on Artificial Intelligence and Machine Learning. 169.

16. Zhang, C., Benz, P., Imtiaz, T. & Kweon, I. S. Understanding Adversarial Examples From the Mutual Influence of Images and Perturbations. in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 14509–14518 (2020). doi:10.1109/CVPR42600.2020.01453.

17. Biggio, B. & Roli, F. Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition* **84**, 317–331 (2018).