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

## IEEE

[1] E. Tabassi, K. J. Burns, M. Hadjimichael, A. D. Molina-Markham, and J. T. Sexton, “A taxonomy and terminology of adversarial machine learning,” preprint, Oct. 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 (accessed May 28, 2022).

[3] R. R. Wiyatno, A. Xu, O. Dia, and A. de Berker, “Adversarial Examples in Modern Machine Learning: A Review.” arXiv, Nov. 15, 2019. Accessed: May 28, 2022. [Online]. Available: http://arxiv.org/abs/1911.05268

[4] X. Yuan, P. He, Q. Zhu, and X. Li, “Adversarial Examples: Attacks and Defenses for Deep Learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2805–2824, Sep. 2019, doi: 10.1109/TNNLS.2018.2886017.

[5] L. Huang, A. D. Joseph, B. Nelson, B. I. P. Rubinstein, and J. D. Tygar, “Adversarial machine learning,” in *Proceedings of the 4th ACM workshop on Security and artificial intelligence*, New York, NY, USA, Oct. 2011, pp. 43–58. 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 (accessed May 27, 2022).

[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*, Accessed: May 27, 2022. [Online]. Available: https://dlnext.acm.org/doi/abs/10.1002/wics.1511

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

[9] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples.” arXiv, Mar. 20, 2015. Accessed: May 27, 2022. [Online]. Available: http://arxiv.org/abs/1412.6572

[10] Z. Zhao, D. Dua, and S. Singh, “Generating Natural Adversarial Examples.” arXiv, Feb. 23, 2018. Accessed: May 28, 2022. [Online]. Available: http://arxiv.org/abs/1710.11342

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

[12] P. Laskov and R. Lippmann, “Machine learning in adversarial environments,” *Mach Learn*, vol. 81, no. 2, pp. 115–119, Nov. 2010, doi: 10.1007/s10994-010-5207-6.

[13] P. McDaniel, N. Papernot, and Z. B. Celik, “Machine Learning in Adversarial Settings,” *IEEE Security Privacy*, vol. 14, no. 3, pp. 68–72, May 2016, doi: 10.1109/MSP.2016.51.

[14] “Machine learning uncertainties with adversarial neural networks | SpringerLink.” https://link.springer.com/article/10.1140/epjc/s10052-018-6511-8 (accessed May 27, 2022).

[15] R. J. Brachman and P. Stone, “Synthesis Lectures on Artificial Intelligence and Machine Learning,” p. 169.

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

[17] B. Biggio and F. Roli, “Wild patterns: Ten years after the rise of adversarial machine learning,” *Pattern Recognition*, vol. 84, pp. 317–331, Dec. 2018, doi: 10.1016/j.patcog.2018.07.023.