

Assignment 3

G202448018 최은경

- Process

Marabou를 사용하여 외부 신경망 모델(MLP, FashionMNIST)과 데이터셋에 대해 property 기반 검증을 진행하였다.

모델 및 데이터셋은 다음과 같다.

- 2-layer MLP, 28x28 입력, 32 hidden, 10 출력
- FashionMNIST 데이터셋
- PyTorch로 학습 (Test Accuracy: 약 85.7%)
- ONNX로 변환 후 Marabou에 입력 (참고한 링크는 다음과 같다)

■ https://neuralnetworkverification.github.io/Marabou/Examples/0_NNetExample.html

Marabou 실험 과정은 다음과 같다.

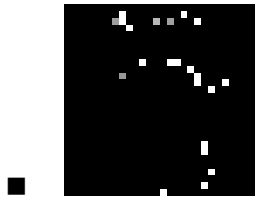
- 입력 변수 : $28 \times 28 = 784$ 차원 벡터 (inputVars), 각 값의 범위는 -1.0 ~ 1.0
- Property 설정 : Marabou에 입력의 bound를 -1.0 ~ 1.0로 제한
- 목적: Marabou SMT solver로 FashionMNIST 2-layer MLP(ONNX 변환 모델)에 대해, output[0] (0번 클래스의 logit)이 [0, 10] 범위를 만족시키는 입력(이미지)을 찾기
- Solver 실행: network.solve() 수행, SAT/UNSAT 여부 및 solution vector 반환

결과 및 해석은 다음과 같다.

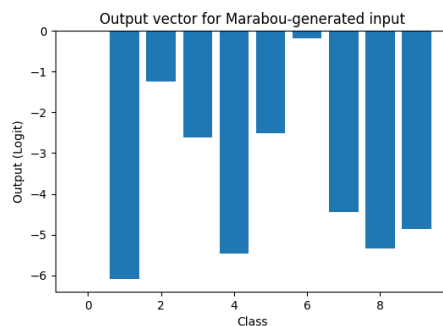
- Marabou 반환값: SAT 반환. 즉, 주어진 property를 만족하는 입력이 존재함
- 반환 input/output 예시화면:

```
input 773 = -1.0
input 774 = -1.0
input 775 = -1.0
input 776 = -1.0
input 777 = -1.0
input 778 = -1.0
input 779 = -1.0
input 780 = -1.0
input 781 = -1.0
input 782 = -1.0
input 783 = -1.0
output 0 = 0.0
output 1 = -6.095442857755011
output 2 = -1.2401168603417156
output 3 = -2.6120612319620244
output 4 = -5.467636288522809
output 5 = -2.511991783493337
output 6 = -0.1973189713889866
output 7 = -4.437740165167481
output 8 = -5.335954050118001
output 9 = -4.870006003492127
["sat", {0: -1.0, 1: -1.0, 2: -1.0, 3: -1.0, 4: -1.0, 5: -1.0, 6: -1.0, 7: -1.0, 8: -1.0, 9: -1.0,
10: -1.0, 11: -1.0, 12: -1.0, 13: -1.0, 14: -1.0, 15: -1.0, 16: -1.0, 17: -1.0, 18: -1.0, 19: -1.0,
20: -1.0, 21: -1.0, 22: -1.0, 23: -1.0, 24: -1.0, 25: -1.0, 26: -1.0, 27: -1.0, 28: -1.0, 29: -1.0,
30: -1.0, 31: -1.0, 32: -1.0, 33: -1.0, 34: -1.0, 35: -1.0, 36: 1.0, 37: -1.0, 38: -1.0, 39: -1.0,
40: -1.0, 41: -1.0, 42: -1.0, 43: -1.0, 44: -1.0, 45: 1.0, 46: -1.0, 47: -1.0, 48: -1.0, 49: -1.0,
50: -1.0, 51: -1.0, 52: -1.0, 53: -1.0, 54: -1.0, 55: -1.0, 56: -1.0, 57: -1.0, 58: -1.0, 59: -1.0,
60: -1.0, 61: -1.0, 62: -1.0, 63: 0.23149714779437922, 64: 1.0, 65: -1.0, 66: -1.0, 67: -1.0, 68:
-1.0, 69: 0.4421504854347666, 70: -1.0, 71: 0.26546305273022397, 72: -1.0, 73: -1.0, 74: -1.0, 75:
1.0, 76: -1.0, 77: -1.0, 78: -1.0, 79: -1.0, 80: -1.0, 81: -1.0, 82: -1.0, 83: -1.0, 84: -1.0, 85:
-1.0, 86: -1.0, 87: -1.0, 88: -1.0, 89: -1.0, 90: -1.0, 91: -1.0, 92: -1.0, 93: 1.0, 94: -1.0, 95:
-1.0, 96: -1.0, 97: -1.0, 98: -1.0, 99: -1.0, 100: -1.0, 101: -1.0, 102: -1.0, 103: -1.0, 104: -1.0,
105: -1.0, 106: -1.0, 107: -1.0, 108: -1.0, 109: -1.0, 110: -1.0, 111: -1.0, 112: -1.0, 113: -1.0,
114: -1.0, 115: -1.0, 116: -1.0, 117: -1.0, 118: -1.0, 119: -1.0, 120: -1.0, 121: -1.0, 122: -1.0,
123: -1.0, 124: -1.0, 125: -1.0, 126: -1.0, 127: -1.0, 128: -1.0, 129: -1.0, 130: -1.0, 131: -1.0,
132: -1.0, 133: -1.0, 134: -1.0, 135: -1.0, 136: -1.0, 137: -1.0, 138: -1.0, 139: -1.0, 140: -1.0,
141: -1.0, 142: -1.0, 143: -1.0, 144: -1.0, 145: -1.0, 146: -1.0, 147: -1.0, 148: -1.0, 149: -1.0,
150: -1.0, 151: -1.0, 152: -1.0, 153: -1.0, 154: -1.0, 155: -1.0, 156: -1.0, 157: -1.0, 158: -1.0,
159: -1.0, 160: -1.0, 161: -1.0, 162: -1.0, 163: -1.0, 164: -1.0, 165: -1.0, 166: -1.0, 167: -1.0,
168: -1.0, 169: -1.0, 170: -1.0, 171: -1.0, 172: -1.0, 173: -1.0, 174: -1.0, 175: -1.0, 176: -1.0,
177: -1.0, 178: -1.0, 179: -1.0, 180: -1.0, 181: -1.0, 182: -1.0, 183: -1.0, 184: -1.0, 185: -1.0,
186: -1.0, 187: -1.0, 188: -1.0, 189: -1.0, 190: -1.0, 191: -1.0, 192: -1.0, 193: -1.0, 194: -1.0,
195: -1.0, 196: -1.0, 197: -1.0, 198: -1.0, 199: -1.0, 200: -1.0, 201: -1.0, 202: -1.0, 203: -1.0,
204: -1.0, 205: -1.0, 206: -1.0, 207: -1.0, 208: -1.0, 209: -1.0, 210: -1.0, 211: -1.0, 212: -1.0,
213: -1.0, 214: -1.0, 215: -1.0, 216: -1.0, 217: -1.0, 218: -1.0, 219: -1.0, 220: -1.0, 221: -1.0,
222: -1.0, 223: -1.0, 224: -1.0, 225: -1.0, 226: -1.0, 227: -1.0, 228: -1.0, 229: -1.0, 230: -1.0,
231: -1.0, 232: -1.0, 233: -1.0, 234: -1.0, 235: -1.0, 236: -1.0, 237: -1.0, 238: -1.0, 239: -1.0,
240: -1.0, 241: -1.0, 242: -1.0, 243: -1.0, 244: -1.0, 245: -1.0, 246: -1.0, 247: -1.0, 248: -1.0,
249: -1.0, 250: -1.0, 251: -1.0, 252: -1.0, 253: -1.0, 254: -1.0, 255: -1.0, 256: -1.0, 257: -1.0,
258: -1.0, 259: -1.0, 260: -1.0, 261: -1.0, 262: -1.0, 263: -1.0, 264: -1.0, 265: -1.0, 266: -1.0,
267: -1.0, 268: -1.0, 269: -1.0, 270: -1.0, 271: -1.0, 272: -1.0, 273: -1.0, 274: -1.0, 275: -1.0,
276: -1.0, 277: -1.0, 278: -1.0, 279: -1.0, 280: -1.0, 281: -1.0, 282: -1.0, 283: -1.0, 284: -1.0,
285: -1.0, 286: -1.0, 287: -1.0, 288: -1.0, 289: -1.0, 290: -1.0, 291: -1.0, 292: -1.0, 293: -1.0,
294: -1.0, 295: -1.0, 296: -1.0, 297: -1.0, 298: -1.0, 299: -1.0, 300: -1.0, 301: -1.0, 302: -1.0,
303: -1.0, 304: -1.0, 305: -1.0, 306: -1.0, 307: -1.0, 308: -1.0, 309: -1.0, 310: -1.0, 311: -1.0,
312: -1.0, 313: -1.0, 314: -1.0, 315: -1.0, 316: -1.0, 317: -1.0, 318: -1.0, 319: -1.0, 320: -1.0,
321: -1.0, 322: -1.0, 323: -1.0, 324: -1.0, 325: -1.0, 326: -1.0, 327: -1.0, 328: -1.0, 329: -1.0,
330: -1.0, 331: -1.0, 332: -1.0, 333: -1.0, 334: -1.0, 335: -1.0, 336: -1.0, 337: -1.0, 338: -1.0,
339: -1.0, 340: -1.0, 341: -1.0, 342: -1.0, 343: -1.0, 344: -1.0, 345: -1.0, 346: -1.0, 347: -1.0,
348: -1.0, 349: -1.0, 350: -1.0, 351: -1.0, 352: -1.0, 353: -1.0, 354: -1.0, 355: -1.0, 356: -1.0,
357: -1.0, 358: -1.0, 359: -1.0, 360: -1.0, 361: -1.0, 362: -1.0, 363: -1.0, 364: -1.0, 365: -1.0,
366: -1.0, 367: -1.0, 368: -1.0, 369: -1.0, 370: -1.0, 371: -1.0, 372: -1.0, 373: -1.0, 374: -1.0,
375: -1.0, 376: -1.0, 377: -1.0, 378: -1.0, 379: -1.0, 380: -1.0, 381: -1.0, 382: -1.0, 383: -1.0,
384: -1.0, 385: -1.0, 386: -1.0, 387: -1.0, 388: -1.0, 389: -1.0, 390: -1.0, 391: -1.0, 392: -1.0,
393: -1.0, 394: -1.0, 395: -1.0, 396: -1.0, 397: -1.0, 398: -1.0, 399: -1.0, 400: -1.0, 401: -1.0,
402: -1.0, 403: -1.0, 404: -1.0, 405: -1.0, 406: -1.0, 407: -1.0, 408: -1.0, 409: -1.0, 410: -1.0,
411: -1.0, 412: -1.0, 413: -1.0, 414: -1.0, 415: -1.0, 416: -1.0, 417: -1.0, 418: -1.0, 419: -1.0,
420: -1.0, 421: -1.0, 422: -1.0, 423: -1.0, 424: -1.0, 425: -1.0, 426: -1.0, 427: -1.0, 428: -1.0,
429: -1.0, 430: -1.0, 431: -1.0, 432: -1.0, 433: -1.0, 434: -1.0, 435: -1.0, 436: -1.0, 437: -1.0,
438: -1.0, 439: -1.0, 440: -1.0, 441: -1.0, 442: -1.0, 443: -1.0, 444: -1.0, 445: -1.0, 446: -1.0,
447: -1.0, 448: -1.0, 449: -1.0, 450: -1.0, 451: -1.0, 452: -1.0, 453: -1.0, 454: -1.0, 455: -1.0,
456: -1.0, 457: -1.0, 458: -1.0, 459: -1.0, 460: -1.0, 461: -1.0, 462: -1.0, 463: -1.0, 464: -1.0,
465: -1.0, 466: -1.0, 467: -1.0, 468: -1.0, 469: -1.0, 470: -1.0, 471: -1.0, 472: -1.0, 473: -1.0,
474: -1.0, 475: -1.0, 476: -1.0, 477: -1.0, 478: -1.0, 479: -1.0, 480: -1.0, 481: -1.0, 482: -1.0,
483: -1.0, 484: -1.0, 485: -1.0, 486: -1.0, 487: -1.0, 488: -1.0, 489: -1.0, 490: -1.0, 491: -1.0,
492: -1.0, 493: -1.0, 494: -1.0, 495: -1.0, 496: -1.0, 497: -1.0, 498: -1.0, 499: -1.0, 500: -1.0,
501: -1.0, 502: -1.0, 503: -1.0, 504: -1.0, 505: -1.0, 506: -1.0, 507: -1.0, 508: -1.0, 509: -1.0,
510: -1.0, 511: -1.0, 512: -1.0, 513: -1.0, 514: -1.0, 515: -1.0, 516: -1.0, 517: -1.0, 518: -1.0,
519: -1.0, 520: -1.0, 521: -1.0, 522: -1.0, 523: -1.0, 524: -1.0, 525: -1.0, 526: -1.0, 527: -1.0,
528: -1.0, 529: -1.0, 530: -1.0, 531: -1.0, 532: -1.0, 533: -1.0, 534: -1.0, 535: -1.0, 536: -1.0,
537: -1.0, 538: -1.0, 539: -1.0, 540: -1.0, 541: -1.0, 542: -1.0, 543: -1.0, 544: -1.0, 545: -1.0,
546: -1.0, 547: -1.0, 548: -1.0, 549: -1.0, 550: -1.0, 551: -1.0, 552: -1.0, 553: -1.0, 554: -1.0,
555: -1.0, 556: -1.0, 557: -1.0, 558: -1.0, 559: -1.0, 560: -1.0, 561: -1.0, 562: -1.0, 563: -1.0,
564: -1.0, 565: -1.0, 566: -1.0, 567: -1.0, 568: -1.0, 569: -1.0, 570: -1.0, 571: -1.0, 572: -1.0,
573: -1.0, 574: -1.0, 575: -1.0, 576: -1.0, 577: -1.0, 578: -1.0, 579: -1.0, 580: -1.0, 581: -1.0,
582: -1.0, 583: -1.0, 584: -1.0, 585: -1.0, 586: -1.0, 587: -1.0, 588: -1.0, 589: -1.0, 590: -1.0,
591: -1.0, 592: -1.0, 593: -1.0, 594: -1.0, 595: -1.0, 596: -1.0, 597: -1.0, 598: -1.0, 599: -1.0,
600: -1.0, 601: -1.0, 602: -1.0, 603: -1.0, 604: -1.0, 605: -1.0, 606: -1.0, 607: -1.0, 608: -1.0,
609: -1.0, 610: -1.0, 611: -1.0, 612: -1.0, 613: -1.0, 614: -1.0, 615: -1.0, 616: -1.0, 617: -1.0,
618: -1.0, 619: -1.0, 620: -1.0, 621: -1.0, 622: -1.0, 623: -1.0, 624: -1.0, 625: -1.0, 626: -1.0,
627: -1.0, 628: -1.0, 629: -1.0, 630: -1.0, 631: -1.0, 632: -1.0, 633: -1.0, 634: -1.0, 635: -1.0,
636: -1.0, 637: -1.0, 638: -1.0, 639: -1.0, 640: -1.0, 641: -1.0, 642: -1.0, 643: -1.0, 644: -1.0,
645: -1.0, 646: -1.0, 647: -1.0, 648: -1.0, 649: -1.0, 650: -1.0, 651: -1.0, 652: -1.0, 653: -1.0,
654: -1.0, 655: -1.0, 656: -1.0, 657: -1.0, 658: -1.0, 659: -1.0, 660: -1.0, 661: -1.0, 662: -1.0,
663: -1.0, 664: -1.0, 665: -1.0, 666: -1.0, 667: -1.0, 668: -1.0, 669: -1.0, 670: -1.0, 671: -1.0,
672: -1.0, 673: -1.0, 674: -1.0, 675: -1.0, 676: -1.0, 677: -1.0, 678: -1.0, 679: -1.0, 680: -1.0,
681: -1.0, 682: -1.0, 683: -1.0, 684: -1.0, 685: -1.0, 686: -1.0, 687: -1.0, 688: -1.0, 689: -1.0,
690: -1.0, 691: -1.0, 692: -1.0, 693: -1.0, 694: -1.0, 695: -1.0, 696: -1.0, 697: -1.0, 698: -1.0,
699: -1.0, 700: -1.0, 701: -1.0, 702: -1.0, 703: -1.0, 704: -1.0, 705: -1.0, 706: -1.0, 707: -1.0,
708: -1.0, 709: -1.0, 710: -1.0, 711: -1.0, 712: -1.0, 713: -1.0, 714: -1.0, 715: -1.0, 716: -1.0,
717: -1.0, 718: -1.0, 719: -1.0, 720: -1.0, 721: -1.0, 722: -1.0, 723: -1.0, 724: -1.0, 725: -1.0,
726: -1.0, 727: -1.0, 728: -1.0, 729: -1.0, 730: -1.0, 731: -1.0, 732: -1.0, 733: -1.0, 734: -1.0,
735: -1.0, 736: -1.0, 737: -1.0, 738: -1.0, 739: -1.0, 740: -1.0, 741: -1.0, 742: -1.0, 743: -1.0,
744: -1.0, 745: -1.0, 746: -1.0, 747: -1.0, 748: -1.0, 749: -1.0, 750: -1.0, 751: -1.0, 752: -1.0,
753: -1.0, 754: -1.0, 755: -1.0, 756: -1.0, 757: -1.0, 758: -1.0, 759: -1.0, 760: -1.0, 761: -1.0,
762: -1.0, 763: -1.0, 764: -1.0, 765: -1.0, 766: -1.0, 767: -1.0, 768: -1.0, 769: -1.0, 770: -1.0,
771: -1.0, 772: -1.0, 773: -1.0, 774: -1.0, 775: -1.0, 776: -1.0, 777: -1.0, 778: -1.0, 779: -1.0,
780: -1.0, 781: -1.0, 782: -1.0, 783: -1.0, 784: -1.0, 785: -1.0, 786: -1.0, 787: -1.0, 788: -1.0,
789: -1.0, 790: -1.0, 791: -1.0, 792: -1.0, 793: -1.0, 794: -1.0, 795: -1.0, 796: -1.0, 797: -1.0,
798: -1.0, 799: -1.0, 800: -1.0, 801: -1.0, 802: -1.0, 803: -1.0, 804: -1.0, 805: -1.0, 806: -1.0,
807: -1.0, 808: -1.0, 809: -1.0, 810: -1.0, 811: -1.0, 812: -1.0, 813: -1.0, 814: -1.0, 815: -1.0,
816: -1.0, 817: -1.0, 818: -1.0, 819: -1.0, 820: -1.0, 821: -1.0, 822: -1.0, 823: -1.0, 824: -1.0,
825: -1.0, 826: -1.0, 827: -1.0, 828: -1.0, 829: -1.0, 830: -1.0, 831: -1.0, 832: -1.0, 833: -1.0,
834: -1.0, 835: -1.0, 836: -1.0, 837: -1.0, 838: -1.0, 839: -1.0, 840: -1.0, 841: -1.0, 842: -1.0,
843: -1.0, 844: -1.0, 845: -1.0, 846: -1.0, 847: -1.0, 848: -1.0, 849: -1.0, 850: -1.0, 851: -1.0,
852: -1.0, 853: -1.0, 854: -1.0, 855: -1.0, 856: -1.0, 857: -1.0, 858: -1.0, 859: -1.0, 860: -1.0,
861: -1.0, 862: -1.0, 863: -1.0, 864: -1.0, 865: -1.0, 866: -1.0, 867: -1.0, 868: -1.0, 869: -1.0,
870: -1.0, 871: -1.0, 872: -1.0, 873: -1.0, 874: -1.0, 875: -1.0, 876: -1.0, 877: -1.0, 878: -1.0,
879: -1.0, 880: -1.0, 881: -1.0, 882: -1.0, 883: -1.0, 884: -1.0, 885: -1.0, 886: -1.0, 887: -1.0,
888: -1.0, 889: -1.0, 890: -1.0, 891: -1.0, 892: -1.0, 893: -1.0, 894: -1.0, 895: -1.0, 896: -1.0,
897: -1.0, 898: -1.0, 899: -1.0, 900: -1.0, 901: -1.0, 902: -1.0, 903: -1.0, 904: -1.0, 905: -1.0,
906: -1.0, 907: -1.0, 908: -1.0, 909: -1.0, 910: -1.0, 911: -1.0, 912: -1.0, 913: -1.0, 914: -1.0,
915: -1.0, 916: -1.0, 917: -1.0, 918: -1.0, 919: -1.0, 920: -1.0, 921: -1.0, 922: -1.0, 923: -1.0,
924: -1.0, 925: -1.0, 926: -1.0, 927: -1.0, 928: -1.0, 929: -1.0, 930: -1.0, 931: -1.0, 932: -1.0,
933: -1.0, 934: -1.0, 935: -1.0, 936: -1.0, 937: -1.0, 938: -1.0, 939: -1.0, 940: -1.0, 941: -1.0,
942: -1.0, 943: -1.0, 944: -1.0, 945: -1.0, 946: -1.0, 947: -1.0, 948: -1.0, 949: -1.0, 950: -1.0,
951: -1.0, 952: -1.0, 953: -1.0, 954: -1.0, 955: -1.0, 956: -1.0, 957: -1.0, 958: -1.0, 959: -1.0,
960: -1.0, 961: -1.0, 962: -1.0, 963: -1.0, 964: -1.0, 965: -1.0, 966: -1.0, 967: -1.0, 968: -1.0,
969: -1.0, 970: -1.0, 971: -1.0, 972: -1.0, 973: -1.0, 974: -1.0, 975: -1.0, 976: -1.0, 977: -1.0,
978: -1.0, 979: -1.0, 980: -1.0, 981: -1.0, 982: -1.0, 983: -1.0, 984: -1.0, 985: -1.0, 986: -1.0,
987: -1.0, 988: -1.0, 989: -1.0, 990: -1.0, 991: -1.0, 992: -1.0, 993: -1.0, 994: -1.0, 995: -1.0,
996: -1.0, 997: -1.0, 998: -1.0, 999: -1.0, 1000: -1.0, 1001: -1.0, 1002: -1.0, 1003: -1.0, 1004: -1.0,
1005: -1.0, 1006: -1.0, 1007: -1.0, 1008: -1.0, 1009: -1.0, 1010: -1.0, 1011: -1.0, 1012: -1.0, 1013: -1.0,
1014: -1.0, 1015: -1.0, 1016: -1.0, 1017: -1.0, 1018: -1.0, 1019: -1.0, 1020: -1.0, 1021: -1.0, 1022: -1.0,
1023: -1.0, 1024: -1.0, 1025: -1.0, 1026: -1.0, 1027: -1.0, 1028: -1.0, 1029: -1.0, 1030: -1.0, 1031: -1.0,
1032: -1.0, 1033: -1.0, 1034: -1.0, 1035: -1.0, 1036: -1.0, 1037: -1.0, 1038: -1.0, 1039: -1.0, 1040: -1.0,
1041: -1.0, 1042: -1.0, 1043: -1.0, 1044: -1.0, 1045: -1.0, 1046: -1.0, 1047: -1.0, 1048: -1.0, 1049: -1.0,
1050: -1.0, 1051: -1.0, 1052: -1.0, 1053: -1.0, 1054: -1.0, 1055: -1.0, 1056: -1.0, 1057: -1.0, 1058: -1.0,
1059: -1.0, 1060: -1.0, 1061: -1.0, 1062: -1.0, 1063: -1.0, 1064: -1.0, 1065: -1.0, 1066: -1.0, 1067: -1.0,
1068: -1.0, 1069: -1.0, 1070: -1.0, 1071: -1.0, 1072: -1.0, 1073: -1.0, 1074: -1.0, 1075: -1.0, 1076: -1.0,
1077: -1.0, 1078: -1.0, 1079: -1.0, 1080: -1.0, 1081: -1.0, 1082: -1.0, 1083: -1.0, 1084: -1.0, 1085: -1.0,
1086: -1.0, 1087: -1.0, 1088: -1.0, 1089: -1.0, 1090: -1.0, 1091: -1.0, 1092: -1.0, 1093: -1.0, 1094: -1.0,
1095: -1.0, 1096: -1.0, 1097: -1.0, 1098: -1.0, 1099: -1.0, 1100: -1.0, 1101: -1.0, 1102: -1.0, 1103: -1.0,
1104: -1.0, 1105: -1.0, 1106: -1.0, 1107: -1.0, 1108: -1.0, 1109: -1.0, 1110: -1.0, 1111: -1.0, 1112: -1.0,
1113: -1.0, 1114: -1.0, 1115: -1.0, 1116: -1.0, 1117: -1.0, 111
```

- Marabou가 찾은 입력 벡터(784차원): 설정해놓은 제약(property)을 만족하는 모델의 입력. 예를 들어, output 0의 logit이 0.0이 나오도록 하는 입력을 찾으라는 property를 줬다면, 실제로 그런 입력이 존재함을 보여주는 것으로 파악
- input 벡터의 특성: 대다수 값이 -1.0 (최소 bound), 일부 값이 1.0 또는 0~1 사이의 값으로 보임.
- 실제 이미지로 복원하면 사람 눈에는 전혀 의미 없는 sparse한 패턴 (실험 이미지 첨부)



- output 해석: output[0]=0.0, 나머지는 -6~-1 사이의 음수(logit 값). 해당 입력이 주어졌을 때 모델은 output 0에 대한 logit을 0.0, 나머지 클래스는 모두 음수로 예측. 즉, 해당 입력에서는 0번 클래스를 가장 높은 logit으로 분류한다는 의미



결론은 다음과 같다.

Marabou를 이용해 property 기반 검증에 성공했다. 모델의 property (예: 특정 logit이 가장 높도록 하는 입력) 만족 입력이 존재함을 확인하였으며 해당 입력은 실제 데이터셋에 없는, 인위적인 패턴의 입력임을 관찰했다.