Kaggle Project
Applied AI Project 5

Description:

The algorithm that I have used in my project is called Gradient Boosting Regression. Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

Running Instructions:

To run the program, the user must have Python 3.5.1 and Anaconda 2.5.0 installed. Given below are the instructions to install Python and Anaconda.

Windows Anaconda install:

In your browser download the Anaconda installer [link https://www.continuum.io/downloads] for Windows, then double click the .exe file and follow the instructions on the screen. If unsure about any setting, simply accept the defaults as they all can be changed later.
NOTE: When finished, a new terminal window will open. If not, click Start - Run - Command Prompt.

Windows Anaconda update:

Open a terminal window with Start - Run - Command Prompt, navigate to the anaconda folder, then type conda update conda.

Windows Anaconda Uninstall:

Go to Control Panel, click “Add or remove Program,” select “Python 2.7 (Miniconda)” and click Remove Program.

After the installation is complete, perform the following steps:

1. Navigate to the folder where the Kaggle.py file is on command prompt and execute: **python Kaggle.py**
2. If using IDLE, Visual Studio or likewise, simply run the program.

Output of the program:

Loading data ...

Data loaded (read 188318 rows)

Preparing encoder ...

Encoder prepared.

Beginning training ...

Transforming training data ...

Training data transformed into 1154 columns.

Iter	Train Loss	Remaining Time
1	8212333.1643	45.62s
2	7906979.6792	44.47s
3	7650599.3873	44.33s
4	7433524.0565	43.59s
5	7251367.0155	43.31s
6	7096263.1531	42.74s
7	6966161.4017	42.19s
8	6854339.4840	41.73s
9	6747505.2564	41.29s
10	6650160.5136	40.82s
20	6030074.0285	35.99s
30	5678236.2573	31.44s
40	5439404.5479	26.89s
50	5260035.6625	22.37s
60	5122144.3084	17.88s
70	5014176.0178	13.54s
80	4923832.5348	9.01s
90	4848782.6507	4.50s
100	4782346.1858	0.00s

Training complete!

Initiating testing of model ...

Chunk #0: MAE = 1502.0116404674247

Chunk #1: MAE = 1483.9522650853162

Chunk #2: MAE = 1375.765357547586

Chunk #3: MAE = 1440.4492111338247

Chunk #4: MAE = 1425.452110040435

Chunk #5: MAE = 1359.1721223796224

Chunk #6: MAE = 1388.7698132090686

Chunk #7: MAE = 1410.1214109758057

Chunk #8: MAE = 1318.3618389872727

Chunk #9: MAE = 1391.1688518062538

Chunk #10: MAE = 1451.2239167720973

Chunk #11: MAE = 1433.8812867555455

Chunk #12: MAE = 1407.2643989830622

Chunk #13: MAE = 1401.4037035981314

Chunk #14: MAE = 1380.2756173794276

Chunk #15: MAE = 1521.514851726362

Chunk #16: MAE = 1395.3761052561667

Chunk #17: MAE = 1307.2384122868377

Chunk #18: MAE = 1407.1073923360834

Chunk #19: MAE = 1364.1123847113686

Chunk #20: MAE = 1356.9245852981505

Chunk #21: MAE = 1529.2938288007867

Chunk #22: MAE = 1399.3171642196974

Chunk #23: MAE = 1409.7787440227794

Chunk #24: MAE = 1447.445069619885

Chunk #25: MAE = 1385.8335144992084

Chunk #26: MAE = 1404.5208291303918

Chunk #27: MAE = 1407.5124767045463

Chunk #28: MAE = 1420.152667136691

Chunk #29: MAE = 1472.256427355153

Chunk #30: MAE = 1506.6056384806657

Chunk #31: MAE = 1446.8123052854864

Chunk #32: MAE = 1396.5403775900666

Chunk #33: MAE = 1600.8718130346374

Chunk #34: MAE = 1418.9945278626658

Chunk #35: MAE = 1445.4810454243413

Chunk #36: MAE = 1418.7771841659776

Chunk #37: MAE = 1491.582902099296

Chunk #38: MAE = 1376.1714351224869

Chunk #39: MAE = 1348.3195421742791

Chunk #40: MAE = 1390.242726367232

Chunk #41: MAE = 1341.0526138210264

Chunk #42: MAE = 1430.6434156590435

Chunk #43: MAE = 1351.5112789203677

Chunk #44: MAE = 1434.9088854676822

Chunk #45: MAE = 1386.9940468601903

Chunk #46: MAE = 1442.786258436955

Chunk #47: MAE = 1523.442386175038

Chunk #48: MAE = 1349.9030169192965

Chunk #49: MAE = 1368.8834766772927

Chunk #50: MAE = 1419.9437490151033

Chunk #51: MAE = 1424.336789721028

Chunk #52: MAE = 1469.3707212849354

Chunk #53: MAE = 1471.1729178961762

Chunk #54: MAE = 1310.8177090996542

Chunk #55: MAE = 1477.709200154675

Chunk #56: MAE = 1419.3640658636937

Chunk #57: MAE = 1421.6885240553167

Chunk #58: MAE = 1520.6214045312468

Chunk #59: MAE = 1426.0969844056424

Chunk #60: MAE = 1412.9402888091802

Chunk #61: MAE = 1563.3889375343097

Chunk #62: MAE = 1461.1259285787828

Chunk #63: MAE = 1509.6180422696075

Chunk #64: MAE = 1407.0899638052556

Chunk #65: MAE = 1450.6751012066784

Chunk #66: MAE = 1376.922129038995

Chunk #67: MAE = 1338.9345574664023

Chunk #68: MAE = 1365.3661649443757

Chunk #69: MAE = 1435.60785956034

Chunk #70: MAE = 1511.6371779666292

Chunk #71: MAE = 1380.7679967928987

Chunk #72: MAE = 1458.6428074938221

Chunk #73: MAE = 1408.499304215817

Chunk #74: MAE = 1364.1028217204528

Chunk #75: MAE = 1453.3169408583276

Chunk #76: MAE = 1427.5405683588024

Chunk #77: MAE = 1538.8794092312178

Chunk #78: MAE = 1336.1884800466914

Chunk #79: MAE = 1444.593358125858

Chunk #80: MAE = 1463.6992072004716

Chunk #81: MAE = 1420.92116827863

Chunk #82: MAE = 1463.0273250502212

Chunk #83: MAE = 1377.5512854128349

Chunk #84: MAE = 1451.1802803420478

Chunk #85: MAE = 1380.2586152168349

Chunk #86: MAE = 1422.0486098194465

Chunk #87: MAE = 1422.5200549552483

Chunk #88: MAE = 1414.2021634241255

Chunk #89: MAE = 1437.333964207205

Chunk #90: MAE = 1423.9946393557332

Chunk #91: MAE = 1386.6028704775667

Chunk #92: MAE = 1376.976239844073

Chunk #93: MAE = 1396.4750334590847

Chunk #94: MAE = 1423.2338298593816

Chunk #95: MAE = 1471.4709971879647

Chunk #96: MAE = 1377.7991401804259

Chunk #97: MAE = 1399.3681865794974

Chunk #98: MAE = 1562.4961875076913

Chunk #99: MAE = 1341.3849186057435

Chunk #100: MAE = 1452.1575987216463

Chunk #101: MAE = 1485.2636212415664

Chunk #102: MAE = 1411.4721445508599

Chunk #103: MAE = 1408.1619574216975

Chunk #104: MAE = 1511.294734915249

Chunk #105: MAE = 1477.012939956896

Chunk #106: MAE = 1398.0607614849512

Chunk #107: MAE = 1492.7799606617762

Chunk #108: MAE = 1500.8077129363658

Chunk #109: MAE = 1458.3557204881647

Chunk #110: MAE = 1406.023404540635

Chunk #111: MAE = 1396.9584980084198

Chunk #112: MAE = 1385.508336175642

Chunk #113: MAE = 1435.6844995214842

Chunk #114: MAE = 1357.8655892810896

Chunk #115: MAE = 1445.2136045156667

Chunk #116: MAE = 1356.317893880066

Chunk #117: MAE = 1508.305598506344

Chunk #118: MAE = 1390.5984093647542

Chunk #119: MAE = 1391.5915241488897

Chunk #120: MAE = 1427.1655417021739

Chunk #121: MAE = 1460.5273367418324

Chunk #122: MAE = 1471.678066567217

Chunk #123: MAE = 1504.032185765745

Chunk #124: MAE = 1461.9623730344165

Chunk #125: MAE = 1529.1441078637135

Chunk #126: MAE = 1354.1630368532535

Chunk #127: MAE = 1382.7242434794846

Chunk #128: MAE = 1400.2663417684864

Chunk #129: MAE = 1389.677894549237

Chunk #130: MAE = 1360.4745878499375

Chunk #131: MAE = 1496.1495754348402

Chunk #132: MAE = 1415.2247864498695

Chunk #133: MAE = 1532.0053670568786

Chunk #134: MAE = 1412.4001400905943

Chunk #135: MAE = 1288.821749327829

Chunk #136: MAE = 1424.3368016175336

Chunk #137: MAE = 1421.7587203230655

Chunk #138: MAE = 1511.9836336918531

Chunk #139: MAE = 1383.0612582537403

Chunk #140: MAE = 1421.6449951034117

Chunk #141: MAE = 1393.9043057243873

Chunk #142: MAE = 1428.3482072649153

Chunk #143: MAE = 1424.3821160018238

Chunk #144: MAE = 1392.5825049058212

Chunk #145: MAE = 1448.012598405981

Chunk #146: MAE = 1294.7814402124268

Chunk #147: MAE = 1386.0382803954765

Chunk #148: MAE = 1462.5519013898866

Chunk #149: MAE = 1414.9009996255431

Chunk #150: MAE = 1328.6542270076582

Chunk #151: MAE = 1413.2901995942402

Chunk #152: MAE = 1456.1034679367226

Chunk #153: MAE = 1501.2566013189562

Chunk #154: MAE = 1418.6111281327362

Chunk #155: MAE = 1346.1878969579811

Chunk #156: MAE = 1507.78206605436

Chunk #157: MAE = 1404.7837045651615

Chunk #158: MAE = 1472.6561292121714

Chunk #159: MAE = 1451.6648219530766

Chunk #160: MAE = 1376.27843126986

Chunk #161: MAE = 1468.974960984148

Chunk #162: MAE = 1360.3035411615833

Chunk #163: MAE = 1482.8567098768856

Chunk #164: MAE = 1343.5114659662813

Chunk #165: MAE = 1449.4703296727935

Chunk #166: MAE = 1494.6453356302943

Chunk #167: MAE = 1405.9603036525693

Chunk #168: MAE = 1516.6101523011998

Result MAE = 1425.1563821112857