10/1/2019

Predicting Loan Approvals

A sample model with applications to claims

J.T. Bassett

Agenda

Olive

- What is Python? (2.5 mins)
- Packages (7.5 mins)
- Get into the code!
  - Initial Exploration (10 mins)
  - Missing Values (5 mins)
  - Preprocessing and Building (15 mins)
  - Parameter Adjustment (10 mins)
- Next Steps and Applications (5 mins)
- Q&A (5 mins)

Before we begin

- No time necessary to set up python (use binder link)
- No time for debugging -> use “final” code
- No background -> Focus on **concepts** first, syntax later
- Q&A throughout

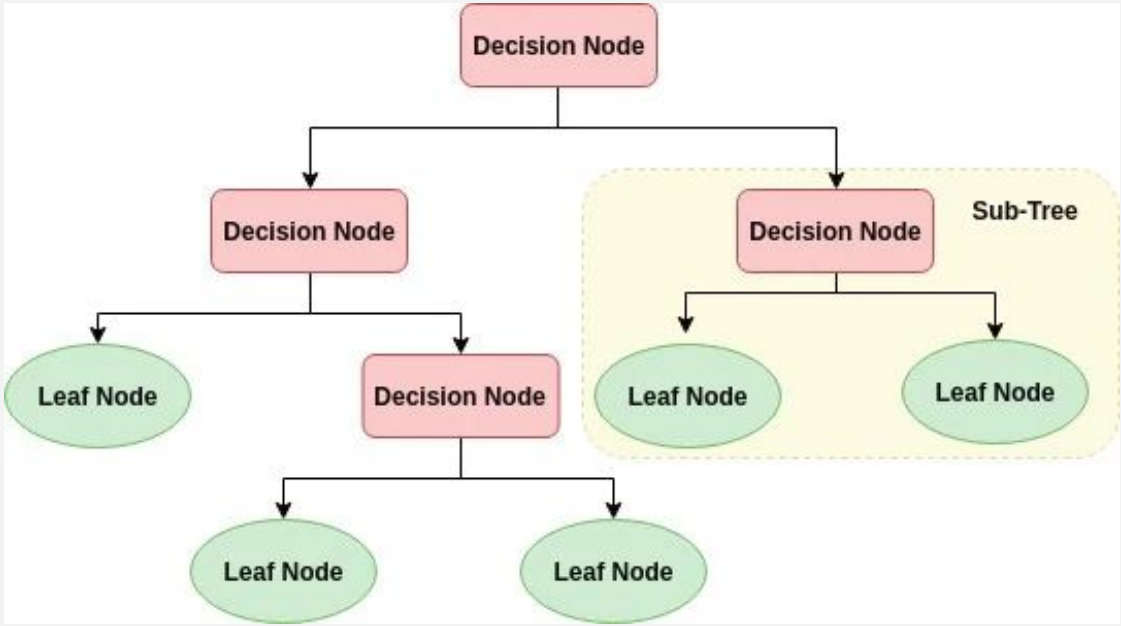
So what is Python?

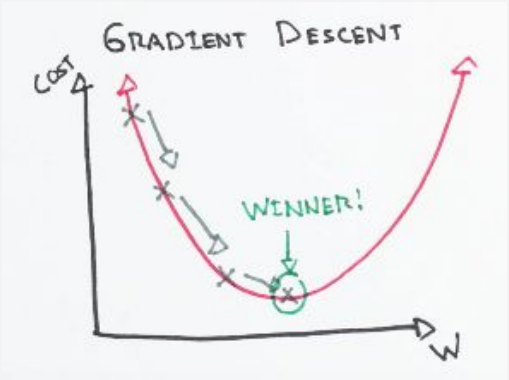
- High-level general-purpose programming language
- The primary programming language used for:
  - Data Manipulation
  - Model Building
- Open source
- Documentation

Python Packages

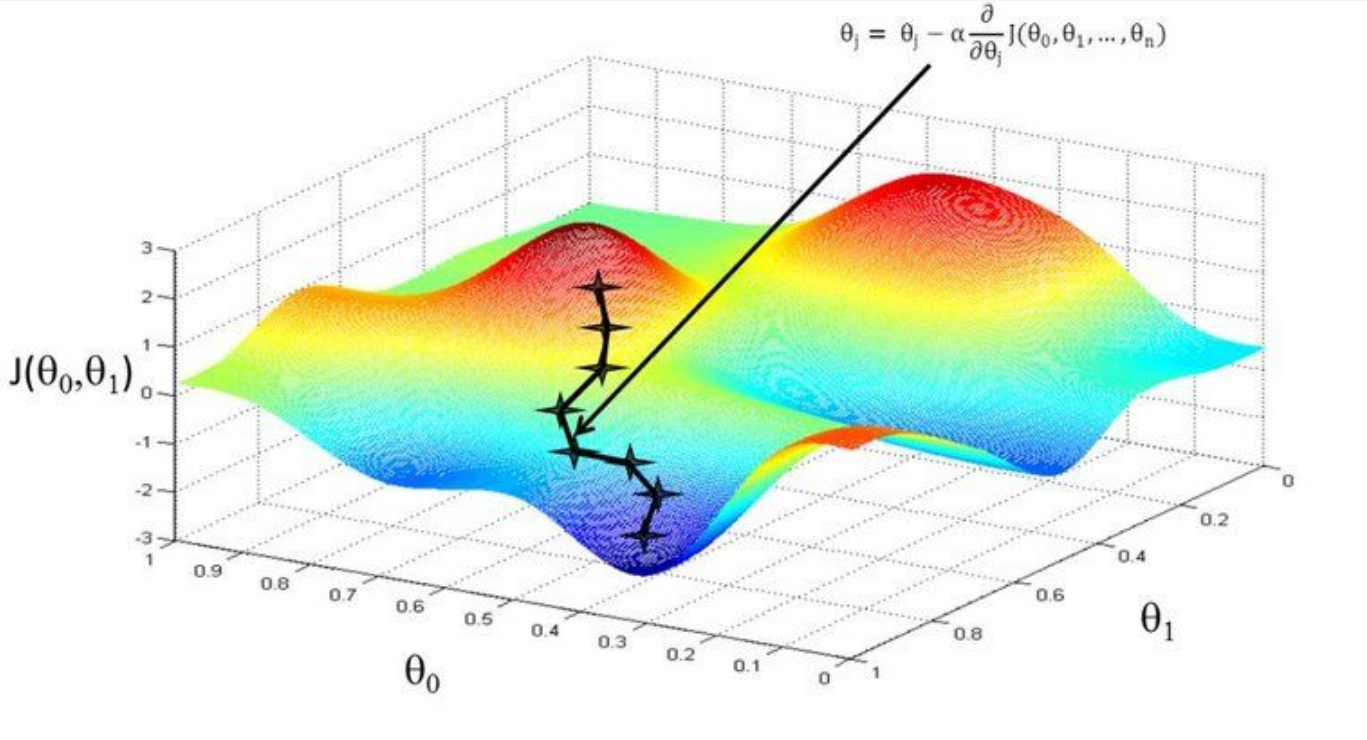
- Pandas
- Numpy
- Matplotlib
- Sklearn
- Xgboost

XGBoost


Decision Node
Decision Node
Decision Node
Sub-Tree
Leaf Node
Decision Node
Leaf Node
Leaf Node
Leaf Node
Leaf Node


GRADIENT DESCENT
COST
WINNER!
W

XGBoost

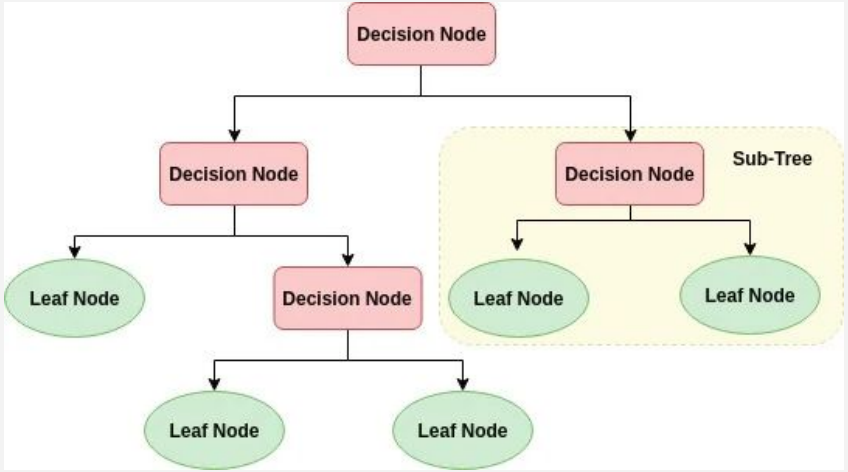

$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1, \ldots, \theta_n)$
$J(\theta_0,\theta_1)$
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ_0
0
0.2
0.4
0.6
0.8
1
θ_1

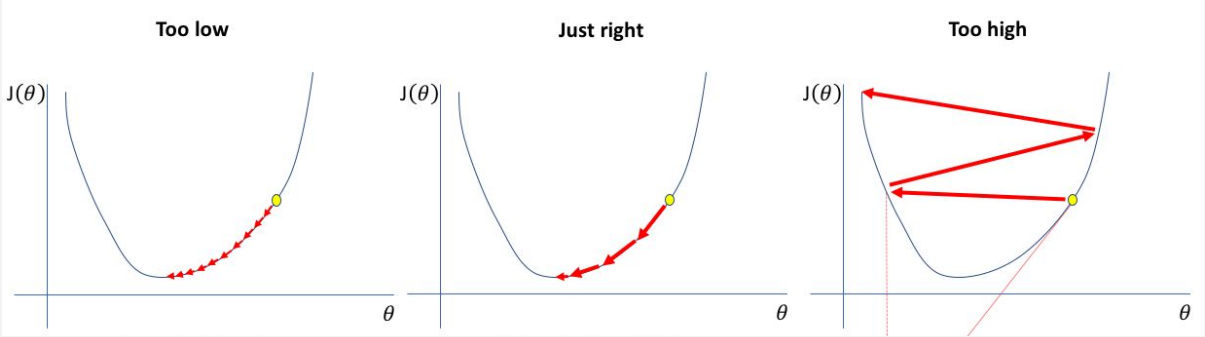
How are we feeling?

Let’s get to the code!

Key Parameters

- Max Depth
- Learning Rate
- Number of Estimators
- Objective
- Number of Jobs

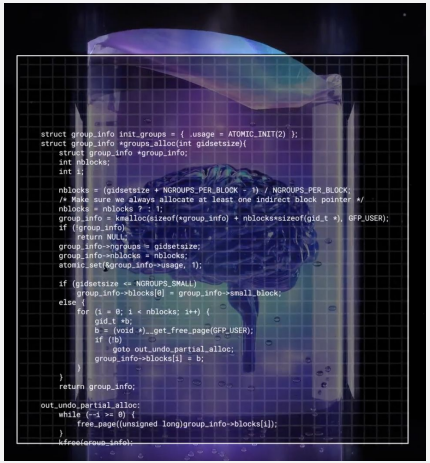




Next Steps

- Continue to adjust parameters
- Try handling nulls differently
- Try different objective (precision v. recall)
- Try different model
- Collect more data

Questions?


struct group_info init_groups = { .usage = ATOMIC_INIT(2) };
struct group_info *groups_alloc(int gidsetsize){
struct group_info *group_info;
int nblocks;
int i;
nblocks = (gidsetsize + NGROUPS_PER_BLOCK - 1) / NGROUPS_PER_BLOCK;
/* Make sure we always allocate at least one indirect block pointer */
nblocks = nblocks ? : 1;
group_info = kmalloc(sizeof(*group_info) + nblocks*sizeof(gid_t *), GFP_USER);
if (!group_info)
return NULL;
group_info->ngroups = gidsetsize;
group_info->nblocks = nblocks;
atomic_set(&group_info->usage, 1);
if (gidsetsize <= NGROUPS_SMALL)
group_info->blocks[0] = group_info->small_block;
else {
for (i = 0; i < nblocks; i++) {
gid_t *b;
b = (void *)__get_free_page(GFP_USER);
if (!b)
goto out_undo_partial_alloc;
group_info->blocks[i] = b;
}
}
return group_info;
out_undo_partial_alloc:
while (--i >= 0) {
free_page((unsigned long)group_info->blocks[i]);
}
kfree(group_info);


THE FUTURE

Sites of interest

- https://xgboost.readthedocs.io/en/latest/parameter.html
- https://machinelearningmastery.com