# Pedigree Yield Predictor

In this contest, our job is to estimate which variety will yield better.

What really matter is the order of the yield, not the absolute value.

However, if we can estimate the yield accurately, then the order will be accurate too.

This is the basic idea of my approach.

Then the problem is :

Given the pedigree of a variety and date,location, how to estimate its yield ?

I made use of the following knowledge:

1. its brothers : all the varieties produced by the same pedigree

We can assume same pedigree will produce similar varieties.

This information could be obtained in “Material.csv.”

1. its parents : the pedigree

I assume a variety is similar to its pedigree

1. its grandparents: the pedigree’s pedigree

Since we’ve assumed a variety is similar to its pedigree, then it’s reasonable to assume a variety is similar to its “grandparent” too.

Then our job is to merge the above information together.

I use two estimators to do the job.

**Estimator1:**





1. bias ,bias(loc,year),bias(x) are estimated with a matrix factorization approach.

See function “learn\_model\_ly\_v”.

1. If some variety v is not seen in training data, we can estimate its bias(v) with help of

its pedigress. I simply assume bias(v) takes the average bias of its pedigree.

See function “smooth\_bias\_mid”

(3)is implemented in function “predict\_y1”

I give more weights to its brother, less weights to parents and grandparent.

**Estimator2:**



Sometimes we have seen v’s brother be tested in the same location and year.

This information could be very useful.

(1)is estimated with a matrix factorization approach.

See function “learn\_model\_mly”

**Final Estimator:**



See function “predict\_rep”

NOTE:

I’ve refactored my code to make it more readable, see **pyield\_final.cpp**

I’ve tested it in my local environment.