Lab7

2023-04-05

```
#Demonstrating convergence in incomes across racial groups using all-race/gender model
#set seed
HUID <- 21519588
set.seed(HUID)

#Demonstrating model across two generations
#Gen 1-2
parents_rank <- 57.9
kids_rank <- 33.31 + 0.351 * parents_rank
kids_rank
```

```
## [1] 53.6329
```

```
#Gen 2-3
parents_rank = kids_rank
kids_rank = 33.31 + 0.351 * parents_rank
kids_rank
```

```
## [1] 52.13515
```

```
#Iterating across multiple generations
generations <- seq(1,7,1)

parents_rank_white = 57.9
parents_rank_black = 32.7

#white gen for loop
for(i in generations){
  kids_rank <- 33.31 + 0.351 * parents_rank_white
  print(paste0("In generation ", i, ", parent_rank = ", parents_rank_white, ", child_rank = ", kids_rank
  parents_rank_white <- kids_rank
}
```

```
## [1] "In generation 1, parent_rank = 57.9, child_rank = 53.6329"
## [1] "In generation 2, parent_rank = 53.6329, child_rank = 52.1351479"
## [1] "In generation 3, parent_rank = 52.1351479, child_rank = 51.6094369129"
## [1] "In generation 4, parent_rank = 51.6094369129, child_rank = 51.4249123564279"
## [1] "In generation 5, parent_rank = 51.4249123564279, child_rank = 51.3601442371062"
## [1] "In generation 6, parent_rank = 51.3601442371062, child_rank = 51.3374106272243"
## [1] "In generation 7, parent_rank = 51.3374106272243, child_rank = 51.3294311301557"
```

```
#black gen for loop
for(i in generations){
  kids_rank <- 33.31 + 0.351 * parents_rank_black
  print(paste0("In generation ", i, ", parent_rank = ", parents_rank_black, ", child_rank = ", kids_rank
  parents_rank_black <- kids_rank
}
```

```
## [1] "In generation 1, parent_rank = 32.7, child_rank = 44.7877"
## [1] "In generation 2, parent_rank = 44.7877, child_rank = 49.0304827"
## [1] "In generation 3, parent_rank = 49.0304827, child_rank = 50.5196994277"
## [1] "In generation 4, parent_rank = 50.5196994277, child_rank = 51.0424144991227"
## [1] "In generation 5, parent_rank = 51.0424144991227, child_rank = 51.2258874891921"
## [1] "In generation 6, parent_rank = 51.2258874891921, child_rank = 51.2902865087064"
## [1] "In generation 7, parent_rank = 51.2902865087064, child_rank = 51.312890564556"
```

Using the all-race/gender model, white and black inter-generational mobility outcomes converge around gen 7 at a rank of about 51.3. But we know that this is incorrect; let's find the steady state prediction for Black and Hispanic children using their respective rank-rank models:

```
#Steady state for Black children
generations <- seq(1,7,1)
parents_rank_black = 32.7

#black gen for loop
for(i in generations){
  kids_rank <- 25.4 + 0.28 * parents_rank_black
  print(paste0("In generation ", i, ", parent_rank = ", parents_rank_black, ", child_rank = ", kids_rank
  parents_rank_black <- kids_rank
}
```

```
## [1] "In generation 1, parent_rank = 32.7, child_rank = 34.556"
## [1] "In generation 2, parent_rank = 34.556, child_rank = 35.07568"
## [1] "In generation 3, parent_rank = 35.07568, child_rank = 35.2211904"
## [1] "In generation 4, parent_rank = 35.2211904, child_rank = 35.261933312"
## [1] "In generation 5, parent_rank = 35.261933312, child_rank = 35.27334132736"
## [1] "In generation 6, parent_rank = 35.27334132736, child_rank = 35.2765355716608"
## [1] "In generation 7, parent_rank = 35.2765355716608, child_rank = 35.277429960065"
```

```
#Steady state for Hispanic children
parents_rank_hisp = 36.17
for(i in generations){
  kids_rank <- 36.14 + 0.26 * parents_rank_hisp
  print(paste0("In generation ", i, ", parent_rank = ", parents_rank_hisp, ", child_rank = ", kids_rank
  parents_rank_hisp <- kids_rank
}
```

```
## [1] "In generation 1, parent_rank = 36.17, child_rank = 45.5442"
## [1] "In generation 2, parent_rank = 45.5442, child_rank = 47.981492"
## [1] "In generation 3, parent_rank = 47.981492, child_rank = 48.61518792"
## [1] "In generation 4, parent_rank = 48.61518792, child_rank = 48.7799488592"
## [1] "In generation 5, parent_rank = 48.7799488592, child_rank = 48.822786703392"
## [1] "In generation 6, parent_rank = 48.822786703392, child_rank = 48.8339245428819"
## [1] "In generation 7, parent_rank = 48.8339245428819, child_rank = 48.8368203811493"
```

The steady state prediction for Black children is around 35.27 and for Hispanic children, 48.83.

Question 2 Cross-validation helps us avoid the overfit problem by addressing the bias-variance tradeoff in machine learning models. More complex models will eventually fit the noise of the training data, which causes the overfit problem. Cross-validation addresses that by evaluating a model's performance with different sets of training data taken from the original dataset. We can cross-validate a portion of the training data to find the optimal model complexity that minimizes RMSPE and over-fitting.