

Select train/test data based on changing seed

Jayan Wijesingha

15 February 2019

Original group propotions

field_id	nn
g1	20
g2	19
g3	35
g4	60
BG	15
BGL	15
GH	15
GHL	15

Code for data partition

```
train_list <- vector("list",5)
test_list <- vector("list",5)
seed_vec <- c(1, 777, 1985, 37213, 104895)

for (k in 1:length(seed_vec)){
  set.seed(seed_vec[k])

  split_sample <- groupdata2::partition(all_df, p = 0.75, cat_col = "field_id",
                                       num_col = NULL, id_col = NULL,
                                       id_aggregation_fn = sum,
                                       force_equal = FALSE, list_out = F)
  train_list[[k]] <- split_sample %>% dplyr::filter(.partitions == 1) %>% data.frame()
  test_list[[k]] <- split_sample %>% dplyr::filter(.partitions == 2) %>% data.frame()
  test_list[[k]]$seed <- rep(seed_vec[k], dim(test_list[[k]])[1])
}
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

Propotion of partioned data

```
for (z in 1:length(seed_vec)) {
  print(sprintf("Seed value: %i", seed_vec[z]))
  print(table(test_list[[z]]$field_id))
  print("\n")
}
```

```
## [1] "Seed value: 1"
##
##  g1  g2  g3  g4  BG BGL  GH GHL
```

```

## 5 5 9 15 4 4 4 4
## [1] "\n"
## [1] "Seed value: 777"
##
## g1 g2 g3 g4 BG BGL GH GHL
## 5 5 9 15 4 4 4 4
## [1] "\n"
## [1] "Seed value: 1985"
##
## g1 g2 g3 g4 BG BGL GH GHL
## 5 5 9 15 4 4 4 4
## [1] "\n"
## [1] "Seed value: 37213"
##
## g1 g2 g3 g4 BG BGL GH GHL
## 5 5 9 15 4 4 4 4
## [1] "\n"
## [1] "Seed value: 104895"
##
## g1 g2 g3 g4 BG BGL GH GHL
## 5 5 9 15 4 4 4 4
## [1] "\n"

```

Summary of the selected test data

seed	row	field	seed	row	field	seed	row	field	seed	row	field	seed	row	field
1	4	g1	777	4	g1	1985	6	g1	37213	10	g1	104895	8	g1
1	9	g1	777	9	g1	1985	7	g1	37213	14	g1	104895	10	g1
1	13	g1	777	13	g1	1985	8	g1	37213	17	g1	104895	12	g1
1	18	g1	777	14	g1	1985	13	g1	37213	19	g1	104895	13	g1
1	19	g1	777	19	g1	1985	16	g1	37213	20	g1	104895	19	g1
1	21	g2	777	30	g2	1985	24	g2	37213	22	g2	104895	22	g2
1	28	g2	777	33	g2	1985	26	g2	37213	26	g2	104895	29	g2
1	30	g2	777	35	g2	1985	29	g2	37213	27	g2	104895	30	g2
1	31	g2	777	37	g2	1985	34	g2	37213	34	g2	104895	33	g2
1	36	g2	777	38	g2	1985	35	g2	37213	38	g2	104895	37	g2
1	41	g3	777	46	g3	1985	40	g3	37213	41	g3	104895	42	g3
1	50	g3	777	49	g3	1985	45	g3	37213	44	g3	104895	47	g3
1	52	g3	777	53	g3	1985	59	g3	37213	54	g3	104895	50	g3
1	56	g3	777	54	g3	1985	61	g3	37213	59	g3	104895	54	g3
1	61	g3	777	55	g3	1985	62	g3	37213	62	g3	104895	57	g3
1	65	g3	777	61	g3	1985	65	g3	37213	63	g3	104895	68	g3
1	67	g3	777	62	g3	1985	66	g3	37213	64	g3	104895	70	g3
1	69	g3	777	65	g3	1985	67	g3	37213	72	g3	104895	71	g3
1	73	g3	777	73	g3	1985	71	g3	37213	74	g3	104895	74	g3
1	76	g4	777	76	g4	1985	80	g4	37213	77	g4	104895	81	g4
1	77	g4	777	84	g4	1985	83	g4	37213	78	g4	104895	86	g4
1	80	g4	777	88	g4	1985	84	g4	37213	79	g4	104895	89	g4
1	85	g4	777	89	g4	1985	85	g4	37213	84	g4	104895	90	g4
1	101	g4	777	93	g4	1985	86	g4	37213	100	g4	104895	100	g4
1	102	g4	777	99	g4	1985	91	g4	37213	102	g4	104895	101	g4
1	111	g4	777	106	g4	1985	93	g4	37213	109	g4	104895	106	g4
1	112	g4	777	107	g4	1985	95	g4	37213	114	g4	104895	107	g4
1	114	g4	777	110	g4	1985	105	g4	37213	120	g4	104895	109	g4
1	119	g4	777	112	g4	1985	108	g4	37213	123	g4	104895	113	g4
1	120	g4	777	119	g4	1985	119	g4	37213	125	g4	104895	118	g4
1	122	g4	777	121	g4	1985	120	g4	37213	128	g4	104895	121	g4
1	123	g4	777	123	g4	1985	123	g4	37213	132	g4	104895	123	g4
1	128	g4	777	126	g4	1985	125	g4	37213	133	g4	104895	130	g4
1	129	g4	777	127	g4	1985	134	g4	37213	134	g4	104895	133	g4
1	135	BG	777	139	BG	1985	139	BG	37213	138	BG	104895	140	BG
1	141	BG	777	140	BG	1985	142	BG	37213	140	BG	104895	141	BG
1	144	BG	777	141	BG	1985	147	BG	37213	142	BG	104895	148	BG
1	148	BG	777	145	BG	1985	148	BG	37213	144	BG	104895	149	BG
1	150	BGL	777	158	BGL	1985	151	BGL	37213	156	BGL	104895	150	BGL
1	159	BGL	777	161	BGL	1985	155	BGL	37213	157	BGL	104895	154	BGL
1	161	BGL	777	163	BGL	1985	156	BGL	37213	159	BGL	104895	156	BGL
1	164	BGL	777	164	BGL	1985	159	BGL	37213	160	BGL	104895	157	BGL
1	165	GH	777	165	GH	1985	166	GH	37213	167	GH	104895	165	GH
1	171	GH	777	167	GH	1985	169	GH	37213	172	GH	104895	170	GH
1	177	GH	777	176	GH	1985	170	GH	37213	174	GH	104895	174	GH
1	179	GH	777	179	GH	1985	174	GH	37213	179	GH	104895	177	GH
1	180	GHL	777	184	GHL	1985	183	GHL	37213	181	GHL	104895	188	GHL
1	188	GHL	777	185	GHL	1985	184	GHL	37213	184	GHL	104895	190	GHL
1	190	GHL	777	189	GHL	1985	185	GHL	37213	186	GHL	104895	191	GHL
1	191	GHL	777	190	GHL	1985	192	GHL	37213	192	GHL	104895	192	GHL