

Logistic Regression: Multinom

Import libraries as usual...

```
library(nnet)
```

Reading the data...

```
train_values <- read.csv(
  "../data/Richters_Predictor_Modeling_Earthquake_Damage_-_Train_Values.csv"
)
train_labels <- read.csv(
  "../data/Richters_Predictor_Modeling_Earthquake_Damage_-_Train_Labels.csv"
)
```

We append the target variable to the full dataset, factor it and relevel it, a necessary step for the multinom algorithm. Relevel simply makes one of the possible categories the “reference”.

```
full_data = train_values
full_data$damage_grade = train_labels$damage_grade
full_data$damage_grade = factor(full_data$damage_grade, levels = c(1, 2, 3), labels=c("low damage", "medium damage", "high damage"))
full_data$damage_grade = relevel(full_data$damage_grade, ref=1)
```

Let us create a function that runs the multinom algorithm, and afterwards calculates the accuracy of the given model. Will take n, which is the number of rows we will use (set up at lower values when exploring so the algorithm does not take lots of time) and formula, in which we specify what variables we want to keep or discard.

```
prediction = function(n, formula){
  #build model
  model = multinom(formula, full_data[1:n,])

  #error calculation
  confMat = table(predict(model), full_data[1:n,]$damage_grade)

  accuracy = sum(diag(confMat))/sum(confMat)

  z = summary(model)$coefficients/summary(model)$standard.errors
  p <- (1 - pnorm(abs(z), 0, 1)) * 2
  print(p)

  return(accuracy)
}
```

This will make formula exploring much easier, since we can now just define n and formula, call the function and keep track of the differences in the accuracy's model.

Attaching the data so we can have access to the variables.

```
attach(full_data)
```

We'll start with 10 000 rows and all variables.

```
n = 10000 #260601
formula = damage_grade ~ .
accuracy1 = prediction(n, formula)
```

```
## # weights: 189 (124 variable)
## initial value 10986.122887
## iter 10 value 9696.205911
## iter 20 value 8493.305728
## iter 30 value 8199.823452
## iter 40 value 7953.133742
## iter 50 value 7876.752849
## iter 60 value 7848.194210
## iter 70 value 7844.909446
## iter 80 value 7843.352040
## iter 90 value 7842.124532
## iter 100 value 7841.249515
## final value 7841.249515
## stopped after 100 iterations
## (Intercept) building_id geo_level_1_id geo_level_2_id
## medium damage 0 0.08041031 0 0.02259624
## almost destroyed 0 0.02491785 0 0.08873523
## geo_level_3_id count_floors_pre_eq age area_percentage
## medium damage 0.008189617 0 0 0
## almost destroyed 0.019761686 0 0 0
## height_percentage land_surface_conditiono
## medium damage 0 0
## almost destroyed 0 0
## land_surface_conditiont foundation_typei foundation_typer
## medium damage 0 0 0
## almost destroyed 0 0 0
## foundation_typeu foundation_typev roof_typeq roof_typex
## medium damage 0 0 0 0
## almost destroyed 0 0 0 0
## ground_floor_typem ground_floor_typev ground_floor_typex
## medium damage 0 0 0
## almost destroyed 0 0 0
## ground_floor_typez other_floor_typeq other_floor_types
## medium damage 0 0 0
## almost destroyed 0 0 0
## other_floor_typex positiono positions positiont
## medium damage 0 0 0 0
## almost destroyed 0 0 0 0
## plan_configurationnc plan_configurationnd plan_configurationf
## medium damage 0 0 0
## almost destroyed 0 0 0
## plan_configurationnm plan_configurationnn plan_configurationo
## medium damage 0 0 0
```

## almost destroyed	0	0	0
##	plan_configurationq	plan_configurations	plan_configurationu
## medium damage	0	0	0
## almost destroyed	0	0	0
##	has_superstructure_adobe_mud		
## medium damage	0		
## almost destroyed	0		
##	has_superstructure_mud_mortar_stone		
## medium damage	0		
## almost destroyed	0		
##	has_superstructure_stone_flag		
## medium damage	0		
## almost destroyed	0		
##	has_superstructure_cement_mortar_stone		
## medium damage	0		
## almost destroyed	0		
##	has_superstructure_mud_mortar_brick		
## medium damage	0		
## almost destroyed	0		
##	has_superstructure_cement_mortar_brick		
## medium damage	0		
## almost destroyed	0		
##	has_superstructure_timber	has_superstructure_bamboo	
## medium damage	0	0	
## almost destroyed	0	0	
##	has_superstructure_rc_non_engineered		
## medium damage	0		
## almost destroyed	0		
##	has_superstructure_rc_engineered	has_superstructure_other	
## medium damage	0	0	
## almost destroyed	0	0	
##	legal_ownership_statusr	legal_ownership_statusv	
## medium damage	0	0	
## almost destroyed	0	0	
##	legal_ownership_statusw	count_families	has_secondary_use
## medium damage	0	0	0
## almost destroyed	0	0	0
##	has_secondary_use_agriculture	has_secondary_use_hotel	
## medium damage	0	0	
## almost destroyed	0	0	
##	has_secondary_use_rental	has_secondary_use_institution	
## medium damage	0	0	
## almost destroyed	0	0	
##	has_secondary_use_school	has_secondary_use_industry	
## medium damage	0	0	
## almost destroyed	0	0	
##	has_secondary_use_health_post	has_secondary_use_gov_office	
## medium damage	0	NaN	
## almost destroyed	0	NaN	
##	has_secondary_use_use_police	has_secondary_use_other	
## medium damage	NaN	0	
## almost destroyed	NaN	0	

```
accuracy1
```

```
## [1] 0.5937
```

We got almost 0.6, which is not very bad, but could be a lot better.

Let's remove the geo levels 2 and 3, since they are too granular to give any information, and also the building id, which provides absolutely no information.

```
formula = damage_grade ~ . -building_id - geo_level_2_id - geo_level_3_id
accuracy2 = prediction(n, formula)
```

```
## # weights: 180 (118 variable)
## initial value 10986.122887
## iter 10 value 9011.647781
## iter 20 value 8561.042569
## iter 30 value 8156.121618
## iter 40 value 7956.128547
## iter 50 value 7871.732758
## iter 60 value 7852.933739
## iter 70 value 7850.337896
## iter 80 value 7849.178637
## iter 90 value 7847.766319
## iter 100 value 7847.176647
## final value 7847.176647
## stopped after 100 iterations
## (Intercept) geo_level_1_id count_floors_pre_eq age
## medium damage 2.544614e-02 3.311509e-10 0.1324426269 0.08981588
## almost destructed 1.833441e-09 5.897890e-02 0.0002594504 0.15485961
## area_percentage height_percentage land_surface_conditiono
## medium damage 0.3067980 0.1899411 0.9959909
## almost destructed 0.8282815 0.2731493 0.5497962
## land_surface_conditiont foundation_typeei foundation_typer
## medium damage 0.2585174 0.4533627 0.05943847
## almost destructed 0.7985651 0.3755110 0.11005181
## foundation_typeeu foundation_typev roof_typeeq roof_typex
## medium damage 0.04256024 0.34281780 5.713552e-03 0.23180421
## almost destructed 0.32882196 0.03074828 1.141854e-06 0.03471077
## ground_floor_typem ground_floor_typev ground_floor_typex
## medium damage 0.2659667 1.819557e-01 0.9676628
## almost destructed 0.8080716 4.414040e-06 0.9139848
## ground_floor_typez other_floor_typeeq other_floor_types
## medium damage 0.7522545 0.002720146 0.005896677
## almost destructed 0.4745477 0.682173746 0.017892756
## other_floor_typex positiono positions positiont
## medium damage 0.06341554 0.9164951 0.1144784 0.359668995
## almost destructed 0.64388351 0.8742566 0.4718908 0.002334726
## plan_configurationnc plan_configurationnd plan_configurationf
## medium damage 0.03282908 8.062414e-04 0
## almost destructed 0.00000000 9.483059e-11 0
## plan_configurationnm plan_configurationnn plan_configurationo
## medium damage 0.8888844 0 0.3290001
## almost destructed 0.0000000 0 0.1343017
```

##	plan_configurationq	plan_configurations	plan_configurationu
## medium damage	1.856697e-03	0.050222304	0.18978461
## almost destroyed	3.469891e-11	0.002388579	0.07088852
##	has_superstructure_adobe_mud		
## medium damage	1.918790e-05		
## almost destroyed	2.137037e-05		
##	has_superstructure_mud_mortar_stone		
## medium damage	5.417888e-14		
## almost destroyed	3.330669e-15		
##	has_superstructure_stone_flag		
## medium damage	2.134782e-03		
## almost destroyed	4.447072e-06		
##	has_superstructure_cement_mortar_stone		
## medium damage	0.44449383		
## almost destroyed	0.04971995		
##	has_superstructure_mud_mortar_brick		
## medium damage	1.994248e-05		
## almost destroyed	3.622453e-03		
##	has_superstructure_cement_mortar_brick		
## medium damage	4.915833e-01		
## almost destroyed	6.967520e-08		
##	has_superstructure_timber	has_superstructure_bamboo	
## medium damage	0.0609408316	0.41717652	
## almost destroyed	0.0000907402	0.03689253	
##	has_superstructure_rc_non_engineered		
## medium damage	0.091070740		
## almost destroyed	0.006595444		
##	has_superstructure_rc_engineered	has_superstructure_other	
## medium damage	0.008646976	0.8880423	
## almost destroyed	0.099878468	0.9313274	
##	legal_ownership_statusr	legal_ownership_statusv	
## medium damage	0.5559514	0.3449941	
## almost destroyed	0.3770801	0.5071857	
##	legal_ownership_statusw	count_families	has_secondary_use
## medium damage	0.6545155	6.440687e-07	0
## almost destroyed	0.1996916	4.021506e-07	0
##	has_secondary_use_agriculture	has_secondary_use_hotel	
## medium damage	0	0	
## almost destroyed	0	0	
##	has_secondary_use_rental	has_secondary_use_institution	
## medium damage	0	0.000000e+00	
## almost destroyed	0	6.619127e-09	
##	has_secondary_use_school	has_secondary_use_industry	
## medium damage	0	1.634426e-11	
## almost destroyed	0	2.533515e-05	
##	has_secondary_use_health_post	has_secondary_use_gov_office	
## medium damage	7.435169e-08	1	
## almost destroyed	0.000000e+00	NaN	
##	has_secondary_use_use_police	has_secondary_use_other	
## medium damage	1	0.1547214	
## almost destroyed	NaN	0.3546700	

accuracy2

```
## [1] 0.5934
```