Warm Up: Machine Learning with a Heart

HOSTED BY DRIVENDATA

Glory!

GLORY!

11
MONTHS LEFT

LEADERBOARD (/COMPETITIONS/54/MACHINE-LEARNING-WITH-A-HEART/LEADERBOARD/)

DATA DOWNLOAD (/COMPETITIONS/54/MACHINE-LEARNING-WITH-A-HEART/DATA/)

SUBMISSIONS (/COMPETITIONS/54/MACHINE-LEARNING-WITH-A-HEART/SUBMISSIONS/)

TEAM (/COMPETITIONS/54/MACHINE-LEARNING-WITH-A-HEART/TEAM/)

DISCUSSION (HTTPS://COMMUNITY.DRIVENDATA.ORG/C/WARM-UP-HEART)   6

OFFICIAL RULES (/COMPETITIONS/54/MACHINE-LEARNING-WITH-A-HEART/RULES/)

SHARE THIS:

Facebook   Twitter   LinkedIn   Email

Problem description

Your goal is to predict the binary class `heart_disease_present` , which represents whether or not a patient has heart disease:

- `0` represents no heart disease present
- `1` represents heart disease present

Features	Performance metric	Submission Format
List of features	Example	Format example
Example of features		

Dataset

There are 14 columns in the dataset, where the `patient_id` column is a unique and random identifier. The remaining 13 features are described in the section below.

- `slope_of_peak_exercise_st_segment` (type: int): the slope of the peak exercise ST segment (https://en.wikipedia.org/wiki/ST_segment), an electrocardiography read out indicating quality of blood flow to the heart
- `thal` (type: categorical): results of thallium stress test (https://www.ucsfbenioffchildrens.org/tests/007201.html) measuring blood flow to the heart, with possible values `normal`, `fixed_defect`, `reversible_defect`
- `resting_blood_pressure` (type: int): resting blood pressure
- `chest_pain_type` (type: int): chest pain type (4 values)
- `num_major_vessels` (type: int): number of major vessels (0-3) colored by flourosopy
- `fasting_blood_sugar_gt_120_mg_per_dl` (type: binary): fasting blood sugar > 120 mg/dl
- `resting_ekg_results` (type: int): resting electrocardiographic results (values 0,1,2)
- `serum_cholesterol_mg_per_dl` (type: int): serum cholestoral in mg/dl
- `oldpeak_eq_st_depression` (type: float): oldpeak = ST depression (https://en.wikipedia.org/wiki/ST_depression) induced by exercise relative to rest, a measure of abnormality in electrocardiograms
- `sex` (type: binary): `0`: female, `1`: male
- `age` (type: int): age in years
- `max_heart_rate_achieved` (type: int): maximum heart rate achieved (beats per minute)
- `exercise_induced_angina` (type: binary): exercise-induced chest pain (`0`: False, `1`: True)

Feature data example

Here's an example of one of the rows in the dataset so that you can see the kinds of values you might expect in the dataset. Some are binary, some are integers, some are floats, and some are categorical. There are no missing values.

field	value
slope_of_peak_exercise_st_segment	2
thal	normal
resting_blood_pressure	125
chest_pain_type	3
num_major_vessels	0
fasting_blood_sugar_gt_120_mg_per_dl	1
resting_ekg_results	2
serum_cholesterol_mg_per_dl	245
oldpeak_eq_st_depression	2.4
sex	1
age	51
max_heart_rate_achieved	166
exercise_induced_angina	0

Performance metric

Performance is evaluated according to binary **log loss**
(https://en.wikipedia.org/wiki/Loss_functions_for_classification#Cross_entropy_loss_(Log_Loss)).

Submission format

The format for the submission file is two columns with the `patient_id` and `heart_disease_present`. This competition uses log loss as its evaluation metric, so the `heart_disease_present` values you should submit are the **probabilities that a patient has heart disease** (not the binary label).

For example, if you predicted...

patient_id	heart_disease_present
olalu7	0.5
z9n6mx	0.5
5k4413	0.5
mrg7q5	0.5
uki4do	0.5
...	...
bwoyg6	0.5
j8i7ve	0.5
t2zn1n	0.5
oxf8kj	0.5
aeiv0y	0.5

Your `.csv` file that you submit would look like:

```
patient_id,heart_disease_present
olalu7,0.5
z9n6mx,0.5
5k4413,0.5
mrg7q5,0.5
uki4do,0.5
kev1sk,0.5
9n6let,0.5
jxmtyg,0.5
...
5bbknr,0.5
hr6pjx,0.5
r4hsar,0.5
4cezdf,0.5
palhcc,0.5
bwoyg6,0.5
j8i7ve,0.5
t2zn1n,0.5
oxf8kj,0.5
aeiv0y,0.5
```

GLORY!

11
MONTHS LEFT

Good luck!

Good luck and enjoy this problem! If you have any questions you can always visit the user forum (http://community.drivendata.org/)!

DRIVENDATA

ABOUT DRIVENDATA

What we do (/about/)

Who we are (http://drivendata.co/#team)

Blog (http://blog.drivendata.org)

LEGAL

Terms of Use (/termsofuse/)

Copyright Policy (/copyrightpolicy/)

Privacy Policy (/privacypolicy/)

WORK WITH US

As a partner (/partners/)

As a competitor (/competitors/)

Join a competition (/competitions/)

CONTACT

info@drivendata.org (mailto:info@drivendata.org)

DrivenData Inc.

1062 Delaware St.

Denver, CO 80204

Copyright © 2018