

1. Install R packages

We need to install the following dependent packages:

Our method:

R version=4.0.3

Hmisc version=0.8

Rbsurv version=2.48.0

Survival version=3.2-7

Biobase version=2.50.0

randomForestSRC version=2.9.3

survivalROC version=1.0.3

Matrix version=1.2-18

Baseline methods:

survivalsvm version=0.0.5

glmnet version=4.0-2

xgboost version=1.2.0.1

randomForestSRC version=2.9.3

2. Data preprocessing

main ▾ XGBLC / Data_preprocessing /

PanJianqiao first commit	
..	
TCGAprocess_clinical.R	first commit
normalize_and_20percent.R	first commit
run_data_preprocessing.R	first commit

The R codes are used for data preprocessing, and the raw datasets can be collected from biological labs or downloaded from public databases, such as TCGA.

3. If you plan to run the R script on one cancer data set

Please prepare these files as follows

cli.txt	2021/7/9 20:11	文本文档
clinic.txt	2018/7/12 20:06	文本文档
name.txt	2021/7/9 19:49	文本文档
result_mean.R	2020/12/31 11:00	R 文件
run.R	2021/7/12 17:49	R 文件
x.test_1.txt	2021/2/5 23:21	文本文档
x.test_2.txt	2021/2/5 23:21	文本文档
x.test_3.txt	2021/2/5 23:22	文本文档
x.test_4.txt	2021/2/5 23:22	文本文档
x.test_5.txt	2021/2/5 23:23	文本文档
x.train_1.txt	2021/2/5 23:21	文本文档
x.train_2.txt	2021/2/5 23:21	文本文档
x.train_3.txt	2021/2/5 23:22	文本文档
x.train_4.txt	2021/2/5 23:22	文本文档
x.train_5.txt	2021/2/5 23:23	文本文档
xgblc.R	2021/6/21 11:53	R 文件

Put xgblc.R file into the data folder, set the R directory to this cancer datasets, and then run all of them.

These files include RNA sequence data and corresponding clinical information

xgblc.R: xgblc model script

run.R: run script

cli.txt: clinical information

x.test and x.train: RNA sequence data for 5-fold cross validation

cli.txt - 记事本

"Tags"	"A3_T"	"A4_N"	"A5_M"	"A6_Stage"	"age_at_initial_pathologic_diagnosis"	"anatomic_neopla
"TCGA.86.8672"	1	1	1	1	1	1
"TCGA.91.7771"	1	1	2	1	2	1
"TCGA.55.8090"	2	1	1	2	3	2
"TCGA.05.4390"	3	1	1	3	4	2
"TCGA.L9.A50W"	4	2	2	4	5	3
"TCGA.97.7553"	5	1	2	2	4	3
"TCGA.55.7284"	1	1	2	1	6	2
"TCGA.78.7149"	6	1	1	5	7	2
"TCGA.55.6642"	3	1	2	3	8	2
"TCGA.55.8513"	1	1	2	1	9	3
"TCGA.97.8177"	7	1	1	3	1	6
"TCGA.49.AAR9"	1	1	2	1	10	2
"TCGA.97.7547"	3	1	2	3	11	4
"TCGA.71.6725"	3	1	1	3	12	7
"TCGA.38.4627"	4	2	1	4	13	3
"TCGA.78.7147"	3	2	1	1	11	4
"TCGA.75.7030"	1	1	1	1	14	2
"TCGA.93.7347"	2	1	2	2	15	2
"TCGA.55.7574"	7	1	1	3	13	8
"TCGA.53.A4EZ"	7	2	2	4	8	3
"TCGA.55.6987"	5	1	1	2	9	2
"TCGA.80.5607"	3	2	1	1	14	1
"TCGA.78.7145"	6	2	3	6	16	2
"TCGA.73.4675"	1	2	1	7	1	3

xtrain.txt - 写字板

"Tags"	TCGA-BP-4346-01	TCGA-BP-4342-01	TCGA-B4-5378-01	TCGA-BP-4329-01	TCGA-R0-5119-01	TCGA-A3-3316-01	TCGA-CJ-4904-01
RAB4B	5.38818349249	2.41784049789	1.63054169191	2.38489363265	1.66645019563	2.1270585037	2.04290478701
C12orf5	3.5383521035	2.21323384805	2.04853038707	2.81646568359	1.94445586404	4.44324584558	5.2191207
RNF44	23.1314492764	10.3419828731	12.0227784781	6.12977465991	10.761896266	12.1146969442	7.10924136901
DNAH3	0.0206034172591	0.0268723700734	0.00757805083287	0.00892363164179	0.0100178020885	0.0166269709239	0.0103085030981
RLB2A	137.13179268	135.033931496	163.131578599	237.250153593	146.015882256	108.6131004	145.20895
ARL8B	17.8420062345	13.1088685763	28.9129526005	23.9378026952	17.5365572418	19.2211999518	22.9790211317
CALB2	0.0851720830909	0.126460864115	0.474306775265	0.130649373029	0.068445503669	0.308348027365	0.222904675149
MFS03	4.53523822145	2.33122706418	2.07550466283	3.40924390823	3.44362786339	1.50089650863	1.70489826841
PICV	6.94253527747	3.66367318984	6.20974435103	6.35514554322	9.9995053496	3.30966198285	4.55637431301
ZNF708	3.10231922829	1.79034493249	1.74706487896	0.930807363271	1.69340473721	1.59491284899	0.9646229
MYADM12	0.0073763251375	0.00584114053669	0.0	0.0	0.0	0.0065403312844	0.0079603617518
PHX	0.130643222219	0.629715940316	1.24528599424	0.527428626514	0.067817535163	0.264082229642	0.211834074653
MOGAT2	0.00876105666783	0.0	0.00716746016811	0.0603472333266	0.0200321724267	0.0	0.00379268047896
PFN2	9.46607235744	8.80650420253	27.9835167748	18.7008651434	3.61596060965	4.0331789884	6.76588181695
RNMD5A	9.67040345166	7.57969197437	13.8335513656	13.7997152142	15.9812375834	9.00135049331	13.451027
RAD23A	39.134563436	34.9880358861	33.347167822	39.690091643	42.8568309147	27.1411669832	40.766252
OR2D2	0.0	0.0	0.0	0.0	0.0	0.0322025999007	0.0
SERPINE1	0.0	0.0	0.0	0.0	0.00424726853421	0.0052938516938	0.0048356176799
SIN3B	9.1842836912	10.0193495574	9.4564525204	4.3466591409	6.21715757295	7.53212117748	5.64009128832
TLL2	0.0244682958002	0.123301032379	0.0243397584106	0.0181978393055	0.0572016649465	0.159363619952	0.0189197897617
CTSA	49.0614651469	29.7987260724	21.8226900117	19.8660656265	29.4250279206	34.348929341	24.7119496405
IL13	106.294371526	59.33160434	6.53568509974	54.8393299519	74.5073282617	34.0225106398	81.0020759281
FTPM4	1.82965473785	1.70873617096	2.52639323706	1.63940582803	2.00179174172	1.78629188614	1.6977942911
NUCB2	4.52922436369	6.54907378221	5.83222630603	11.1375741252	9.08213405234	7.06075395689	7.85088402436

xtest.txt - 写字板

"Tags"	TCGA-B1-5303-01	TCGA-AR-4845-01	TCGA-CS-5396-01	TCGA-QH-A6IX-01	TCGA-HT-7686-01	TCGA-S9-A7R4-01	TCGA-WY-A858-01
RAB4B	5.42346770597	4.11596918404	2.82680785065	5.16052066657	4.1731787574	3.40391698622	6.61887030942
C12orf5	3.99952537798	2.63127295093	8.78260915558	1.74371198219	5.20708889229	3.21967574273	2.7584246
RNF44	6.71404536743	10.8005593971	11.3292941294	14.8023531852	8.69335863371	12.4755692911	8.14066298328
DNAH3	0.0	0.00175454383396	0.00583370369791	0.00112295396913	0.00835897562976	0.00158352530434	0.0425252624532
RLB2A	103.05257491	106.387133784	70.2997750111	52.0110215725	130.766367307	93.8762690839	104.51415
ARL8B	32.2633986707	33.7938368856	36.5937592651	22.5355582327	38.4375336288	32.2533661531	23.0469531951
CALB2	0.677157119573	3.92468155973	0.331374914126	0.338224643756	1.43036509092	2.25183506337	2.30913881373
MFS03	9.98294843989	13.5100706327	11.3170098433	11.2213962677	4.62605291929	12.2672285177	10.4998064458
PICV	4.06513579448	4.99213107023	2.18474428603	2.78693033431	5.04180433143	3.46334786057	3.70543048329
ZNF708	2.8904133463	5.14861622521	4.68713852788	3.57470905692	3.34903121099	6.1566318004	3.4702592
MYADM12	0.0513145216031	0.481520612269	0.0961073156245	0.276941910945	0.246044788289	0.135666151799	0.0481886
PHX	2.79708007191	0.498937400226	0.694847729133	0.161311410775	1.53941791295	0.492966572384	0.474987216599
MOGAT2	0.0	0.00841051166776	0.00317081127762	0.0	0.00405880439193	0.0	0.035998370601
PFN2	97.7306459188	167.427798959	70.4765779277	78.4079895191	129.443217404	165.134220295	85.4062956782
RNMD5A	7.89709672057	8.37277746957	10.0069443138	8.45273142628	8.10587917323	10.9017717574	7.4299640
RAD23A	32.2335835766	29.3669859943	30.9852584365	64.8715092718	29.0851038969	33.5390727804	42.039499
OR2D2	0.0	0.0	0.0	0.0	0.022371960348	0.0	0.0
SERPINE1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SIN3B	4.46310972793	10.5797687179	7.1279537672	14.1134979003	4.49030103376	9.18251139443	7.66529446169
TLL2	0.265273871603	0.181508045785	0.0499133387584	0.402580664368	0.193736080181	0.0588098870218	0.148223097313
CTSA	14.4863678529	10.596106496	20.5488123043	11.2398734224	19.7226461279	11.5205982906	18.4184544111
IL13	1.43745537505	1.98517691682	0.580743620656	2.6652154308	0.523508612596	1.93544105513	6.18629894276
FTPM4	1.5239289189	2.84516542548	2.45220757004	1.92214135938	2.27382439766	2.7141297965	1.39830129052
NUCB2	3.3393239715	3.58173210253	6.86328393312	5.51477885093	1.979694815	4.44406945076	3.89514480486

4. If you plan to run R script on several cancer data sets together

Place the file as follows

名称	日期	类型
 BRCA	2021/7/11 10:30	文件夹
 CESC	2021/7/11 10:30	文件夹
 HNSC	2021/7/11 10:30	文件夹
 KIRC	2021/7/11 10:30	文件夹
 run.R	2021/7/10 2:31	R 文件
 xgbhc.R	2021/7/10 2:31	R 文件

xgbhc.R: xgbhc model script

run.R: run script

Modify the corresponding directories of run.R and xgbhc.R, for example:

Setwd("File path")

All programs can be downloaded from <https://github.com/lab319/XGBLC>