TAE WK 9 - Bagging & Random Forest Jones 62m 9.1 Boot strapping A resorphy nethod wild to quantity the uncertainty associated in a Houstical learning method. (Other rescripting method: cross-validation) GOOT: EMABLISH emphical dintibusion functions for a wide great range of scotions. -- How to without It show the . We obtain diffinct dota pets by repeatedly sompliny observations thin the original dotates in replacement · Each bootstrap larakes is created by soupling it replacement, f is the same size as our original demoter. of supplied the of a new on so so so such such 25 3; 9 will - Telminology

• Z*1: fill buoling data not the subject to the sub bootstop exmose for x 2 2000 1856 on 16 20 B: NO. Of d. Theren boutstrap doublets. Handab enw $SE_B(a) = \sqrt{B-1} \frac{B}{(a^{AB}-a^{AB})^2}$ of butting of that $SE_B(a) = \sqrt{B-1} \frac{B}{(a^{AB}-a^{AB})^2}$ (- whe it can be applied a dre within on they down with It can be used to tackle the blas-variance trade-off of some HOA. 3t/car learning nethols Louch as Derson tree) 8.2 Bagging (Bootstrap oggregation) Intuition: Derbin Tree Suffer from high variance. Csay If we split the transfor days into 2 pars our random, 4 fix a decision tree to both halves the rekus would be gutte ditt) (For to shirt and rodity cotton) Method: - Take repeated souples from the training doctaret: (Generate B diff bootstrapped troining deservers) in 12 mossing Entire process - Tran model for (x) in the 6-th buothapped training demons 1) colled Crepean for all B donastas.) Bagging - We average all the predictions as follows - I have the

(bag (x) = 6 } f*b(x)

(take moscinity were for classiffication tree)

1

POP bazic

Date:

NO ·

8.3 Dua-Of-Bay (OOB) Error EHMCHM

TAE Wik 9- Barring & Rodon Firth

We use 000 Ever Estimation to text the error of the bagged model

- To obtain a single president for the its observation, we awage president responses (for regression) or mojerity vote (for classification)
 - i. this leads to a single oak predicting for the i'm observation.
- An DOD prediction can be obtained in this way for each of the in observations, from which we compute the owners! OOD 145E1 classificantly error
 - We B is sofficiently large we now a Lecent alternative to cross-calling)

tell land they disting took

8.4 Rondom Forest

Rondom forest builds from the idea of considering a random scape of m

prelictors from the full det of predictors P

- this is to overcome the public to DT bogging where trees that to be correlated.

Idea:

- for each split, we consider only a subject of predictors (P c m)

(Thi) process de correctes or tree)

Algorithm.

- 1. Germane B bootstrapped training demaseds.
- 7. Fur each donors, tom a decision tree (At each Split, UR a subser in of p available pren

8.2 (Coggling (Bort Map Oggrapowin)

3. Autroge the predictions from the B trees
(For clossificants we majority stry)

- · Classificany m = Jp

Hyperparamenes tuning and many on the system of

- · No. of trees, B
- . NO. of predictor used an each Split, m



Jones 62

TAE WK 9 - Ethics of doug onalytics

Ethical issues in dona science

- 1) Bias, discrimination, & exclusion
 Algos & AI errore excussor towards individuals & graps of people.
- algorithmic Profiling

 The modited of pesinolish, can affect the tes colligence principles like democratic of cultural pluvolish of risk sharp in the router of insurance.
- 3) Proceeding massive files while enhancing AI.

 Doma protection lows are noted in the belief that individuals' rights regarding their pessions when pessions were mult be protected.
- The acceptance of the existence of potential bis in domains curated to train, also is of parament inpursame.

Hon data visualilanim is obuited

- 1) Omitty the baseline.

 CNOT HOTTING CA the Origin).
- 2) Monipularly the y-axi)

 (changing the scale of y-axi) to exaggeore or angal trends)
- 3) Cheny Picking Data.
- 4) Using the wong graph
 (wany type of wovolisation).

NO.:	
734 JOSE 1940 J. An TITE - 10 XIII 30	290
5) Zames MARCHES AS INITS	1 -1
\$ 3 44.40 × 54.40 × 10	()
as 6) amitting dang land reserve store 5th 1 2551	
of Conting (Conting)	
7) Extra polarium	
	(0)
P) Unecessor Complexity of the second of the	
dery to what the whole you of exposed traiting I the things	- Vie
Meeting into a first white shoring It.	(3)
1000 portions and out noted in the softe with the soil to the	
1900 में महा प्रदेशको मेहरू मार्ग प्र प्राप्तिकार मार्ग प्रविद्याला.	
	0.37
Dury of Sylvery	(0)
TO OCCUPANTE OF THE PERIOD OF THEM SON I SOUTH TO STANFORD OF	
Bessell + musical 70 , 66,42 84,4	1 375 A
on dona unknohlumen is obuiged.	
Omitting the SOIR line.	
(1/4/10 35 25 (1/1/1)	
By the provided no to the	A Committee of the Comm
1 मिटणाठेश टाम के	
The second of th	
; A.★.¥Y	
CHON PLEAT DEAD.	To the second
May - the willy diely	17
Company to the second of the following of the following	
	P bazic