DALEX

Lung cancer case study

Aleksandra Grudziąż University of Warsaw, MI2 DataLab Munich, 25.05.2018



Data set

	Gender	Smokina	No.mths.	after.givin	na.ur	.smoki	ina In	dustria	al.citv	chronic.br	onchitis	Asthma	Pack.ve	ears.of.sm	okina	Age	
1	М	YES			.9		48		NO		NO	NO			40	63	
2	м	NO					0		NO		NO	NO			0	69	
3	W	YES					0		NO		YES	NO			50	73	
4	М	YES					0		NO		NO	NO			35	59	
5	м	NO					0		NO		NO	NO			40	70	
6	М	NO					0		NO		YES	NO			0	66	
	cancers	s.in.fam	ily Live.	more.than.3	3.y	Lung [Degree	.of.eff	Ficiency	.by.Zubord	Symptoms	s.pain	Symptoms	haemopty:	sis		
1			NO		0	Left				1		YES			NO		
2			NO		0	Left				0		NO			NO		
3			NO			Right				0		NO			NO		
4			NO			Left				0		NO			NO		
5			NO			Right				1		NO			NO		
6			NO		0 F	Right				0		NO			NO		
	Symptom	ns.breat	nlessness	Symptoms.			toms.d	legradat	tion.of	efficiency	Weight.	loss Fe	ature.T	Feature.N		ire.M	Group
1	Symptom	ns.breatl	YES		cough YES	n Sympt S	toms.d	egradat	tion.of	NO	Weight.	loss Fe 0	ature.T 3	Feature.N 1		ure.M O	IIIA
1 2	Symptom	ns.breath	YES NO		cough YES NC	n Sympt S	toms.d	legradat	tion.of	NO	Weight.	loss Fe 0 0	ature.T 3 2	Feature.N 1 1		are.M 0 0	IIIA IIB
1 2 3	Symptom	ns.breath	YES NO NO		COUGH YES NC NC	n Sympt S D	toms.d	legradat	tion.of	NO NO NO	Weight.	0 0 0	ature.T 3 2 2	Feature.N 1 1 0		ure.M 0 0 0	IIIA IIB IB
1 2 3 4	Symptom	ns.breath	YES NO NO		COUGH YES NC NC	n Sympt S D D D	toms.d	legradat	tion.of	NO NO NO		0 0 30	ature.T 3 2 2 2	Feature.N 1 0 0		ure.M 0 0 0	IIIA IIB IB IB
1 2 3 4 5	Symptom	ns.breath	YES NO NO YES		COUGH YES NC NC NC	n Sympt S D D D D	toms.d	legradat	tion.of	NO NO NO NO		0 0 30 0	ature.T 3 2 2 2 2 2	1 1 0 0 2		ure.M 0 0 0 0	IIIA IIB IB IIIA
1 2 3 4 5 6			YES NO NO YES NO		COUGH YES NC NC NC NC	n Symp1 S D D D D D D D				NO NO NO NO NO		0 0 30	ature.T 3 2 2 2 2 3	Feature.N 1 0 0 2 0		ure.M 0 0 0 0 0 0	IIIA IIB IB IB
1 2 3 4 5 6			YES NO NO YES NO	al.diagnos	COUGH YES NC NC NC NC NC	n Symp1 S D D D D D D D	cancer	Tumour	r.volume	NO NO NO NO NO		0 0 30 0	ature.T 3 2 2 2 2 3	1 1 0 0 2		ure.M 0 0 0 0 0	IIIA IIB IB IIIA
1 2 3 4 5 6			YES NO NO YES NO	al.diagnos Missin	Cough YES NC NC NC NC NC Seng	n Symp1 S D D D D D D D	cancer NO	Tumour	^.∨o]ume L.300352	NO NO NO NO NO NO		0 0 30 0	ature.T 3 2 2 2 2 3	1 1 0 0 2		ure.M 0 0 0 0 0	IIIA IIB IB IIIA
1 2 3 4 5 6		l.histopa	YES NO NO YES NO athologic	al.diagnos Missin Missin	cougł YES NC NC NC NC is Se ng	n Symp1 S D D D D D D D	cancer NO NO	Tumour 11	r.volume L.300351	NO NO NO NO NO NO		0 0 30 0	ature.T 3 2 2 2 2 3	1 1 0 0 2		Ire.M 0 0 0 0 0	IIIA IIB IB IIIA
1 2 3 4 5 6		l.histopa Non	YES NO NO YES NO athologic	al.diagnos Missin Missin 11 carcinor	Cough YES NC NC NC NC is Se ng ma	n Symp1 S D D D D D D D	cancer NO NO NO	Tumour 11 11 12	r.volume L.300351 L.831379 ∂.469623	NO NO NO NO NO NO NO		0 0 30 0	ature.T 3 2 2 2 2 3	1 1 0 0 2		Ire.M 0 0 0 0 0 0	IIIA IIB IB IIIA
1 2 3 4 5 6		l.histopa Non	YES NO NO YES NO athologic	al.diagnos Missin Missin 11 carcinor 11 carcinor	Cough YES NC NC NC NC NC NC NC NC MC MC MC MC MC	n Symp1 S D D D D D D D	cancer NC NC NC NC	Tumour 11 11 9 10	r.volume L.300351 L.831379 J.469623	NO NO NO NO NO NO NO 3 3		0 0 30 0	ature.T 3 2 2 2 2 3	1 1 0 0 2		Ire.M 0 0 0 0 0 0	IIIA IIB IB IIIA
1 2 3 4 5 6		l.histopa Non	YES NO NO YES NO athologic	al.diagnos Missin Missin 11 carcinor	cough YES NC NC NC NC NC NC NC NC NC NC NC NC NC	n Symp1 S D D D D D D D	cancer NO NO NO	Tumour 11 11 12 10 10 12	r.volume L.300351 L.831379 ∂.469623	NO NO NO NO NO 2 1 2 3 3 3 5		0 0 30 0	ature.T 3 2 2 2 2 3	1 1 0 0 2		Ire.M 0 0 0 0 0	IIIA IIB IB IIIA

Model

library(mlr)

```
classif_task <- makeClassifTask(id = "rf_model", data = data_model, target = "Live.more.than.3.y")
classif_lrn_rf <- makeLearner("classif.randomForest", predict.type = "prob")</pre>
```

classif_rf <- train(classif_lrn_rf, classif_task, subset=train_index)</pre>



Model and its explainer

library(mlr)

```
classif_task <- makeclassifTask(id = "rf_model", data = data_model, target = "Live.more.than.3.y")
classif_lrn_rf <- makeLearner("classif.randomForest", predict.type = "prob")</pre>
```

```
classif_rf <- train(classif_lrn_rf, classif_task, subset=train_index)</pre>
```

y_test <- as.numeric(as.character(modelTest\$Live.more.than.3.y))</pre>

```
custom_predict_classif <- function(object, newdata) {
    pred <- predict(object, newdata=newdata)
    response <- pred$data[,3]
    return(response)
    }
explainer_classif_rf <- DALEX::explain(classif_rf, data=modelTest, y=y_test,</pre>
```

```
lest, y=y_test,
label= "rf", predict_function = custom_predict_classif)
```



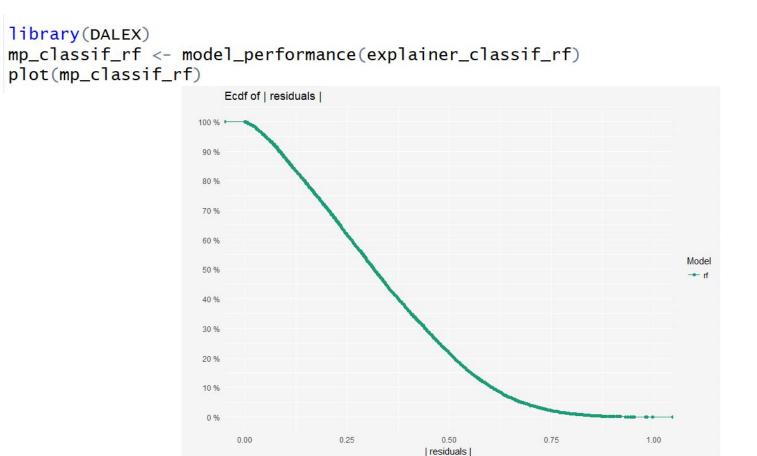
Interpretability with DALEX

Model performance

library(DALEX)
mp_classif_rf <- model_performance(explainer_classif_rf)
plot(mp_classif_rf)</pre>



Model performance





Feature importance

vi_classif_rf <- variable_importance(explainer_classif_rf, loss_function = loss_root_mean_square)
vi_classif_rf</pre>

	variable	dropout_loss	label
1	_baseline_	0.5538452	rf
2	Tumour.volume	0.4593264	rf
3	Age	0.4484891	rf
4	<pre>Initial.histopathological.diagnosis</pre>	0.4254442	rf
5	Years.of.smoking	0.4241094	rf
6	Group	0.4232901	rf
7	Gender	0.4194648	rf
8	Feature.T	0.4109437	rf
9	Degree.of.efficiency.by.Zubord	0.4086999	rf
10	Lung	0.4063517	rf
11	No.mths.after.giving.up.smoking	0.4049651	rf
12	Feature.N	0.4043632	rf
13	Symptoms.cough	0.4024710	rf
14	Symptoms.degradation.of.efficiency	0.4014890	rf
15	Symptoms.pain	0.4009991	rf
16	Smoking	0.4004038	rf
17	Symptoms.haemoptysis	0.3975389	rf
18	Cancers.in.family	0.3971816	rf
19	Symptoms.breathlessness	0.3969818	rf
20	Industrial.city	0.3964980	rf
21	Weight.loss	0.3960647	rf
22	Chronic.bronchitis	0.3937310	rf



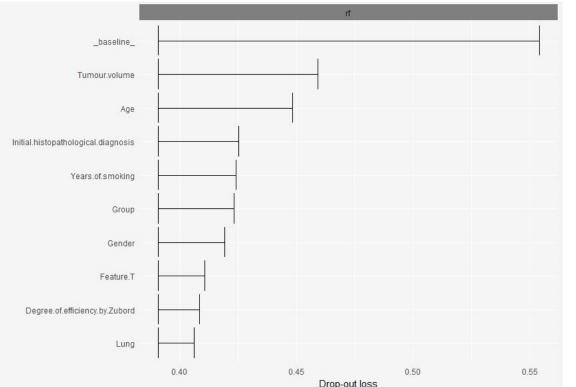
Feature importance

vi_classif_rf <- variable_importance(explainer_classif_rf, loss_function = loss_root_mean_square)
plot(vi_classif_rf)</pre>



Feature importance

vi_classif_rf <- variable_importance(explainer_classif_rf, loss_function = loss_root_mean_square)
plot(vi_classif_rf)</pre>





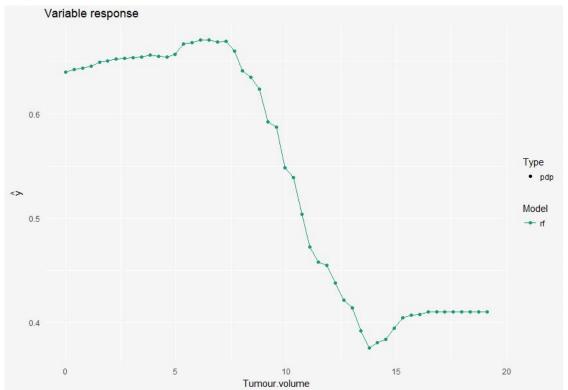
Variable response

pdp_classif_rf <- variable_response(explainer_classif_rf, variable = "Tumour.volume", type = "pdp")
plot(pdp_classif_rf)</pre>



Variable response

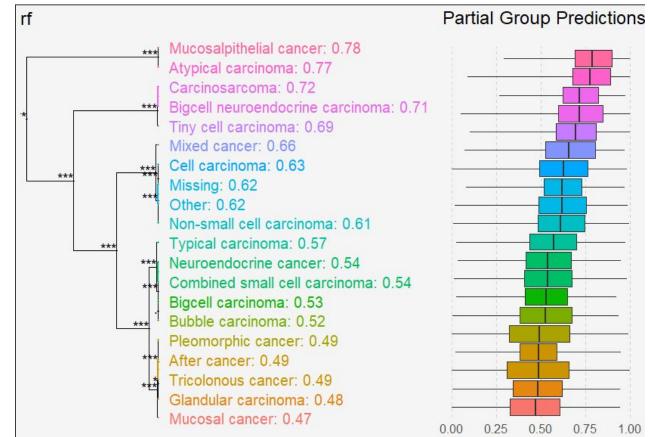
pdp_classif_rf <- variable_response(explainer_classif_rf, variable = "Tumour.volume", type = "pdp")
plot(pdp_classif_rf)</pre>





Merging Path Plot

Merging Path Plot



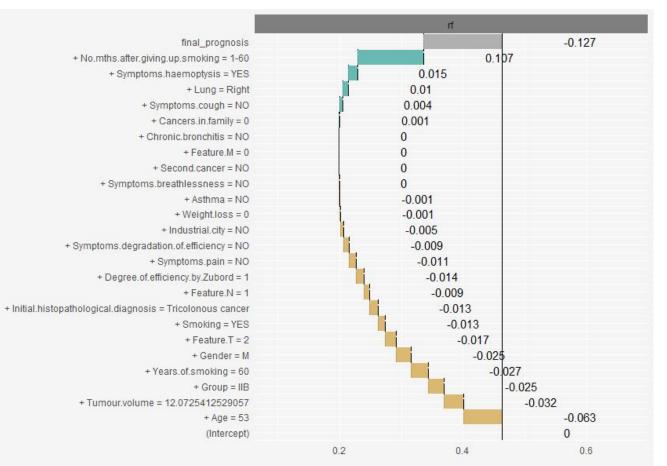


Gender Smoking No.mths.after.giving.up.smoking Industrial.city Chronic.bronchitis Asthma Years.of.smoking Age 2 YES 1 - 60NO NO NO 60 53 M Cancers.in.family Lung Degree.of.efficiency.by.Zubord Symptoms.pain Symptoms.haemoptysis 2 0 Right YES NO Symptoms.breathlessness Symptoms.cough Symptoms.degradation.of.efficiency Weight.loss Feature.T Feature.N 2 NO 0 2 NO 1 NO Feature.M Group Initial.histopathological.diagnosis Second.cancer Tumour.volume 2 Tricolonous cancer 0 IIB NO 12.07254

```
library(breakDown)
library(DALEX)
```

breakDownExplain <- prediction_breakdown(explainer_classif_rf, modelTest[1,])
plot(breakDownExplain)</pre>



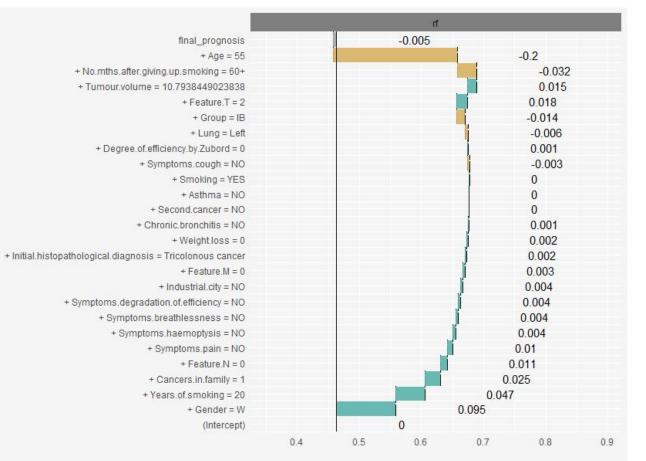


MI

Gender Smoking No.mths.after.giving.up.smoking Industrial.city Chronic.bronchitis Asthma Years.of.smoking 44.1 YES 60 +20 W NO NO NO Age Cancers.in.family Lung Degree.of.efficiency.by.Zubord Symptoms.pain Symptoms.haemoptysis 44.1 55 1 Left NO NO Symptoms.breathlessness Symptoms.cough Symptoms.degradation.of.efficiency Weight.loss Feature.T Feature.N 44.1 0 NO NO NO 2 0 Feature.M Group Initial.histopathological.diagnosis Second.cancer Tumour.volume Tricolonous cancer 44.1 NO 10,79384 0 IB

breakDownExplain <- prediction_breakdown(explainer_classif_rf, modelTest[27,])
plot(breakDownExplain)</pre>







Acknowledgements

We acknowledge the financial support from the NCN Opus grant 2016/21/B/ST6/02176.