

Отчет по лабораторной работе №1 по курсу «Искусственный интеллект»

Выполнила студентка группы 8О-3046 Лаар Марина

Тема: Работа с сервисом Azure-ML

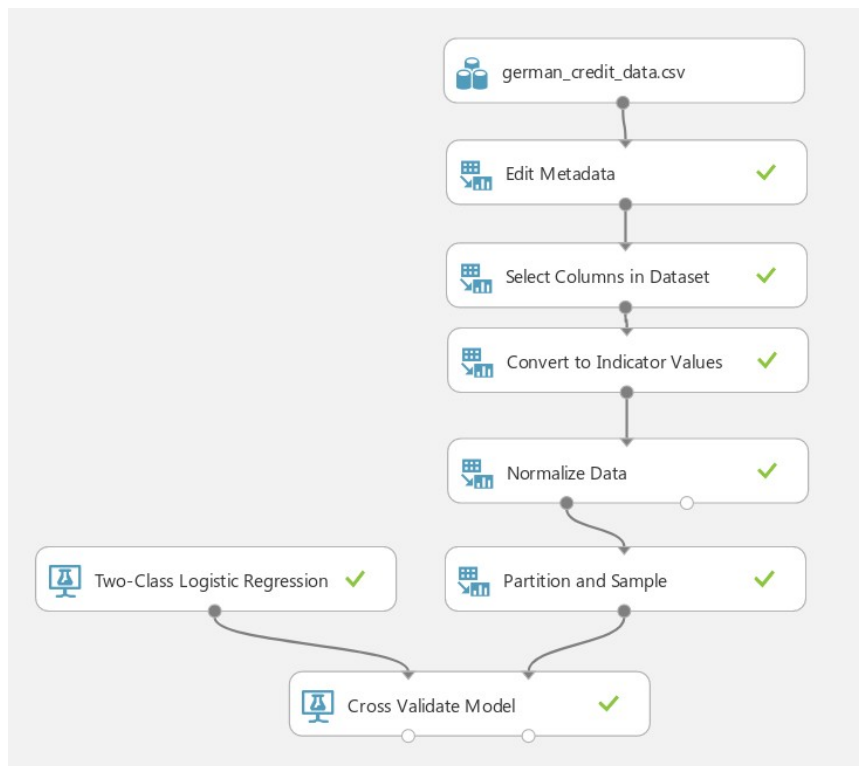
Задание: Познакомиться с веб-сервисом Azure-ML, реализовав полный цикл разработки решения задачи машинного обучения, используя три различных алгоритма, реализованные на данной платформе

Ход работы:

Решение любой задачи машинного обучения начинается с предобработки данных. Существуют некоторые общие подходы, однако довольно часто предобработка является процессом творческим и зависит от конкретной задачи. В данной ЛР можно выделить несколько стадий предобработки (Для каждого алгоритма они будут одни и те же):

- Нормализация числовых признаков
- Dummy-кодирование категориальных признаков (Данная платформа требует, чтобы признаки были явно помечены как категориальные, поэтому также будет использоваться блок Edit Metadata)
- Разбиение датасета на фолды для кросс-валидации

Общая схема решения:



Результат работы блока Convert to Indicator Values:

	Age	Sex-female	Sex-male	Job-0	Job-1	Job-2	Job-3	Housing-free	Housing-own	Housing-rent	Saving accounts-little	Saving accounts-moderate	Saving accounts-NA	Saving accounts-quite rich
view as														
	67	0	1	0	0	1	0	0	1	0	0	0	1	0
	22	1	0	0	0	1	0	0	1	0	1	0	0	0
	49	0	1	0	1	0	0	0	1	0	1	0	0	0
	45	0	1	0	0	1	0	1	0	0	1	0	0	0
	53	0	1	0	0	1	0	1	0	0	1	0	0	0
	35	0	1	0	1	0	0	1	0	0	0	0	1	0
	53	0	1	0	0	1	0	0	1	0	0	0	0	1
	35	0	1	0	0	0	1	0	0	1	1	0	0	0
	61	0	1	0	1	0	0	0	1	0	0	0	0	0
	28	0	1	0	0	0	1	0	1	0	1	0	0	0
	25	1	0	0	0	1	0	0	0	1	1	0	0	0

Результат нормализации:

	Age	Credit amount	Duration
view as			
	0.857143	0.050567	0.029412
	0.053571	0.31369	0.647059
	0.535714	0.101574	0.117647
	0.464286	0.419941	0.558824
	0.607143	0.254209	0.294118
	0.285714	0.484483	0.470588
	0.607143	0.142236	0.294118
	0.285714	0.368548	0.470588
	0.75	0.154561	0.117647
	0.160714	0.274238	0.382353
	0.107143	0.0575	0.117647

Логистическая регрессия:

	Fold Number	Number of examples in fold	Model	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
view as										
	0	200	Logistic Regression	0.775	0.796296	0.914894	0.851485	0.778579	0.503207	17.040324
	1	200	Logistic Regression	0.715	0.732919	0.893939	0.805461	0.764706	0.543524	15.211546
	2	200	Logistic Regression	0.755	0.781065	0.916667	0.84345	0.765129	0.504435	14.928359
	3	200	Logistic Regression	0.71	0.734104	0.913669	0.814103	0.713174	0.560391	8.885622
	4	200	Logistic Regression	0.735	0.782609	0.875	0.82623	0.756572	0.509758	14.030722
	Mean	1000	Logistic Regression	0.738	0.765399	0.902834	0.828146	0.755632	0.524263	14.019315
	Standard Deviation	1000	Logistic Regression	0.027295	0.029709	0.018083	0.019329	0.025013	0.026092	3.071205

Метод опорных векторов:

	Fold Number	Number of examples in fold	Model	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
view as										
	0	200	SVM (Pegasos-Linear)	0.75	0.789809	0.879433	0.832215	0.765717	0.511999	15.590806
	1	200	SVM (Pegasos-Linear)	0.72	0.7375	0.893939	0.808219	0.752785	0.553015	13.731055
	2	200	SVM (Pegasos-Linear)	0.73	0.771084	0.888889	0.825806	0.760045	0.512551	13.559703
	3	200	SVM (Pegasos-Linear)	0.7	0.733728	0.892086	0.805195	0.69796	0.570848	7.185514
	4	200	SVM (Pegasos-Linear)	0.715	0.760479	0.881944	0.81672	0.708953	0.549535	7.322349
	Mean	1000	SVM (Pegasos-Linear)	0.723	0.75852	0.887258	0.817631	0.737092	0.539589	11.477885
	Standard Deviation	1000	SVM (Pegasos-Linear)	0.018574	0.023437	0.006326	0.011441	0.031287	0.026213	3.937632

Нейронная сеть (1 скрытый слой, 128 нейронов)

	Fold Number	Number of examples in fold	Model	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
view as										
	0	200	Binary Neural Network	0.775	0.811688	0.886525	0.847458	0.75009	0.640744	-5.634381
	1	200	Binary Neural Network	0.735	0.739394	0.924242	0.821549	0.740196	0.675465	-5.370853
	2	200	Binary Neural Network	0.685	0.77931	0.784722	0.782007	0.65315	0.845691	-42.623596
	3	200	Binary Neural Network	0.74	0.780645	0.870504	0.823129	0.705154	0.655967	-6.654055
	4	200	Binary Neural Network	0.655	0.755102	0.770833	0.762887	0.668031	0.698606	-17.818124
	Mean	1000	Binary Neural Network	0.718	0.773228	0.847365	0.807406	0.703324	0.703295	-15.620201
	Standard Deviation	1000	Binary Neural Network	0.047645	0.027588	0.066634	0.0342	0.042758	0.082501	15.962235

Выводы:

В данной лабораторной работе я успешно познакомилась с платформой Azure-ML, решила задачу машинного обучения с применением трех разных алгоритмов классификации.