

به نام خدا

نام درس پروژه

عنوان : کاربرد روش های تحلیل چند متغیره در ارزیابی میزان آلودگی هوا

نام و نام خانوادگی دانشجو : حسین فقیهی آغمیونی

استاد راهنما: آقای دکتر حسین باغیثنی

فهرست مطالب

۳.....	مقدمه
۴.....	کاهش بعد داده ها با تحلیل مولفه های اصلی
8.....	یافتن گروه های پنهان در داده ها با تحلیل عاملی
9.....	رده بندی کیفیت هوا: تحلیل ممیزی خطی و درجه دوم
13.....	کدهای استفاده شده

در برخی تحقیقات به دلایل مختلف با حجم زیادی از متغیرها روبرو هستیم. برای تحلیل دقیق تر و رسیدن به نتایج علمی تر و در عین حال عملیاتی تر، پژوهشگر به دنبال کاهش حجم متغیرها و تشکیل ساختار جدیدی برای آنها است و به این منظور از روش تحلیل عاملی استفاده می کنند.

عامل ها متغیر جدیدی هستند که از طریق یک ترکیب خطی از متغیر های مشاهده شده برآورد می شوند. امید این است که با تعداد کمی از عامل ها بتوان تقریباً همه اطلاعاتی که توسط مجموعه بزرگ تری از متغیرها بدست می آید در بر گرفته و در نتیجه توصیف ویژگی های هر مشاهده را ساده ساخت.

تحلیل ممیزی به عنوان یکی از روشهای آماری چند متغیره مطرح میباشد. هدف تحلیل ممیزی تعیین روشی برای تشخیص چند جامعه چند متغیره از یکدیگر و درصدد است تا ترکیب های خطی بین متغیرهای مستقل را که قادرند به بهترین نحو گروه های مشاهدات را از هم جدا کنند، شناسایی کند .

این روش کاربردهای زیادی در مدیریت، علوم پزشکی، اقتصاد، علوم پایه و علوم دیگر دارد.

دو روش تحلیل ممیزی عبارتاند از:

۱) تحلیل ممیزی خطی یا روش فیشر

۲) تحلیل ممیزی درجه دوم

داده های استفاده شده در این پروژه داده های الودگی هوای یکی از شهرهای ایتالیا است که میخواهیم روش های تحلیل چند متغیره را بر روی این داده ها انجام دهیم.

کاهش بعد داده ها با تحلیل مولفه های اصلی

در ابتدا داده هارا در نرم افزار r وارد میکنیم:

```
data=read.table("C:/Users/AH3134/Desktop/uni/PEYVASTE2/wind-quality-white1.txt",header = TRUE)
```

با تابع cor ماتریس همبستگی داده هارا بدست می آوریم:

```
cormatrix=cor(x)
```

ماتریس همبستگی به صورت زیر است:

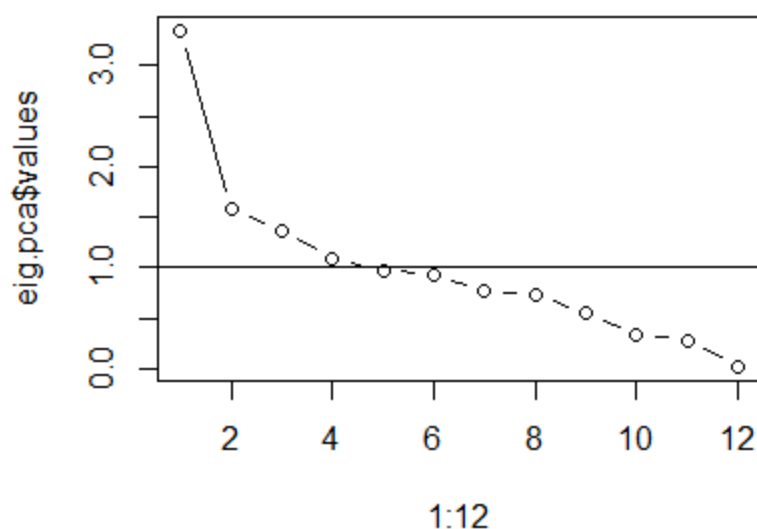
	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides
fixed_acidity	1.00000000	-0.02269729	0.289180698	0.08902070	0.02308564
volatile_acidity	-0.02269729	1.00000000	-0.149471811	0.06428606	0.07051157
citric_acid	0.28918070	-0.14947181	1.000000000	0.09421162	0.11436445
residual_sugar	0.08902070	0.06428606	0.094211624	1.00000000	0.08868454
chlorides	0.02308564	0.07051157	0.114364448	0.08868454	1.00000000
free_sulfur_dioxide	-0.04939586	-0.09701194	0.094077221	0.29909835	0.10139235
total_sulfur_dioxide	0.09106976	0.08926050	0.121130798	0.40143931	0.19891030
density	0.26533101	0.02711385	0.149502571	0.83896645	0.25721132
pH	-0.42585829	-0.03191537	-0.163748211	-0.19413345	-0.09043946
sulphates	-0.01714299	-0.03572815	0.062330940	-0.02666437	0.01676288
alcohol	-0.12088112	0.06771794	-0.075728730	-0.45063122	-0.36018871
quality	-0.11366283	-0.19472297	-0.009209091	-0.09757683	-0.20993441
	free_sulfur_dioxide	total_sulfur_dioxide	density	pH	sulphates
fixed_acidity	-0.0493958591	0.091069756	0.26533101	-0.4258582910	-0.01714299
volatile_acidity	-0.0970119393	0.089260504	0.02711385	-0.0319153683	-0.03572815
citric_acid	0.0940772210	0.121130798	0.14950257	-0.1637482114	0.06233094
residual_sugar	0.2990983537	0.401439311	0.83896645	-0.1941334540	-0.02666437
chlorides	0.1013923521	0.198910300	0.25721132	-0.0904394560	0.01676288
free_sulfur_dioxide	1.0000000000	0.615500965	0.29421041	-0.0006177961	0.05921725
total_sulfur_dioxide	0.6155009650	1.0000000000	0.52988132	0.0023209718	0.13456237
density	0.2942104109	0.529881324	1.00000000	-0.0935914935	0.07449315
pH	-0.0006177961	0.002320972	-0.09359149	1.0000000000	0.15595150
sulphates	0.0592172458	0.134562367	0.07449315	0.1559514973	1.00000000
alcohol	-0.2501039415	-0.448892102	-0.78013762	0.1214320987	-0.01743277
quality	0.0081580671	-0.174737218	-0.30712331	0.0994272457	0.05367788
	alcohol	quality			
fixed_acidity	-0.12088112	-0.113662831			
volatile_acidity	0.06771794	-0.194722969			
citric_acid	-0.07572873	-0.009209091			
residual_sugar	-0.45063122	-0.097576829			
chlorides	-0.36018871	-0.209934411			
free_sulfur_dioxide	-0.25010394	0.008158067			
total_sulfur_dioxide	-0.44889210	-0.174737218			
density	-0.78013762	-0.307123313			
pH	0.12143210	0.099427246			
sulphates	-0.01743277	0.053677877			
alcohol	1.00000000	0.435574715			
quality	0.43557472	1.000000000			

داده هایی که همبستگی زیادی دارند در ماتریس کوواریانس بالا با رنگ های قرمز سبز و زرد مشخص شده اند.

مقادیر ویژه را با تابع eigen مشخص میکنیم :

3.34666886 1.58609030 1.37117206 1.08486150 0.97528337 0.93876295 0.76924825
0.72389378 0.55650103 0.34293339 0.28411542 0.02046911

شش مقدار ویژه ی اول که با رنگ آبی مشخص شده اند مقادیر نسبتاً زیادی دارند و در پیش بینی تحلیل مولفه اصلی میتوان پیش بینی کرد که ابعاد به ۵ یا ۶ کاهش میابد. اندازه ی مقادیر ویژه در نمودار زیر برای درک بهتر رسم شده است که ۴ تا از مقادیر بیشتر از یک و ۲ تا نزدیک به یک است.



برای تعیین بهتر تعداد بعد از تابع princomp که تجزیه طیفی انجام می دهد استفاده میکنیم:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
Sd	1.8293903	1.2594008	1.1709706	1.04156685	0.98756436	0.96889780	0.87706798	0.85081947
PV	0.2788891	0.1321742	0.1142643	0.09040513	0.08127361	0.07823025	0.06410402	0.06032448
CP	0.2788891	0.4110633	0.5253276	0.61573273	0.69700634	0.77523659	0.83934061	0.89966509
	Comp.9	Comp.10	Comp.11	Comp.12				
Standard deviation	0.74598996	0.58560515	0.53302478	0.143070283				
Proportion of Variance	0.04637509	0.02857778	0.02367628	0.001705759				
Cumulative Proportion	0.94604017	0.97461796	0.99829424	1.000000000				

همانطور که مشاهده میشود ۵ مولفه ی اول تا ۷۰ درصد واریانس کل جامعه را توجیه میکند. پس برای تحلیل مولفه های اصلی از ۵ مولفه ی اول فقط استفاده میکنیم.

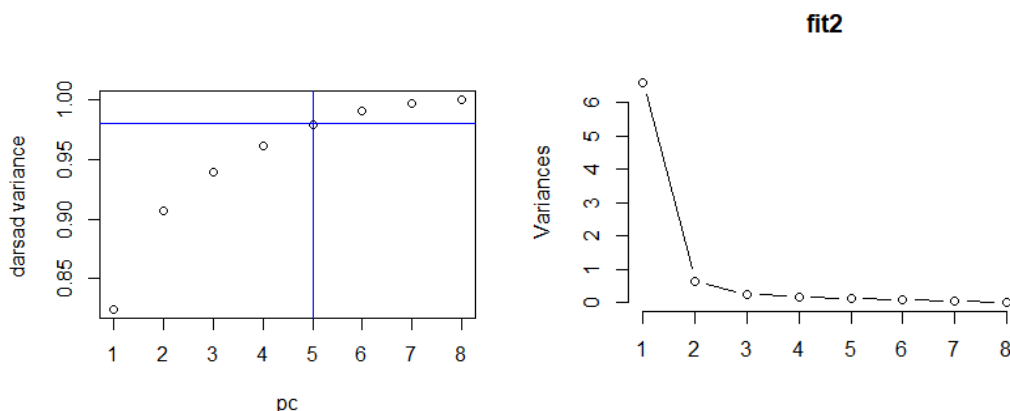
تابع prcomp هم نشان می دهد که انتخاب ۵ مولفه ی اول پیشنهاد خوبی است.

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	2.5683	0.81402	0.5036	0.42327	0.37559	0.29921	0.2350	0.15081

Proportion of Variance 0.8245 0.08283 0.0317 0.02239 0.01763 0.01119 0.0069 0.00284
 Cumulative Proportion 0.8245 0.90733 0.9390 0.96143 0.97907 0.99026 0.9972 1.00000

در این روش ۵ مولفه ی اول ۰/۹۸ درصد واریانس کل جامعه توجیه می کنند. برای درک بهتر نمودار زیر را نگاه کنید.



در نمودار سمت راست مشاهده میشود که مولفه ی اول و دوم سهم بیشتری در تعیین واریانس کل دارند.

نمودار سمت چپ هم نشان دهنده ی این است که ۵ مولفه ی اصلی اول تا ۰/۹۸ درصد واریانس جامعه را توجیه میکنند.

ضرایب بارهای عاملی عبارتند از:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12
fixed_acidity	0.157	0.561	0.207		0.244	0.106	0.224	0.130	0.631	0.201	0.104	0.171
volatile_acidity			-0.525	-0.131	0.703	-0.124	-0.224	-0.230		-0.142	0.270	
citric_acid	0.133	0.289	0.446	0.330		-0.132	-0.120	-0.691	-0.249	-0.106		
residual_sugar	0.406			-0.416		0.290	-0.339	-0.113	-0.177	0.374	-0.180	0.494
chlorides	0.218		-0.215	0.510	-0.178	0.409	-0.552	0.211	0.179	0.236		
free_sulfur_dioxide	0.275	-0.346	0.313	-0.149	0.111	-0.488	0.224	0.129	-0.102	0.327	0.499	
total_sulfur_dioxide	0.390	-0.272	0.125		0.271	-0.272	0.204		0.178	-0.347	-0.644	
density	0.501			-0.104		0.326	-0.123		0.125			-0.761
pH	-0.130	-0.567		0.204	-0.113	0.193		-0.478	0.520	0.184		0.142
sulphates		-0.248	0.227	0.519	0.456	0.480		0.336	-0.237			
alcohol	-0.443		0.159	-0.134	0.309	-0.135				0.575	-0.419	-0.350
quality	-0.227	-0.146	0.489	-0.278			-0.584	0.144	0.300	-0.368	0.146	

حال برای تحلیل عاملی از تابع factanal استفاده میکنیم:

ابتدا مشاهده میکنیم که p-مقدار برابر با ۰/۲۵۴۹ است که نشان میدهد مدل انتخاب شده مدلی خوبی است.

The degrees of freedom for the model is 16 and the fit was 0.2549

برای تحلیل بهتر از درون promax استفاده شده است.

	Factor1	Factor2	Factor3	Factor4	Factor5
ss loadings	1.845	1.639	1.416	1.244	0.744
Proportion var	0.154	0.137	0.118	0.104	0.062
Cumulative Var	0.154	0.290	0.408	0.512	0.574

با استفاده از تحلیل درستیابی ۵ فاکتور اول ۰,۵۷۴ درصد واریانس جامعه را توجیه میکنند.

ماتریس باقی مانده ها به صورت زیر است:

	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides
fixed_acidity	-0.09706585	-0.02195517	-0.021978166	0.27056607	0.08308016
volatile_acidity	-0.02195517	-0.04081220	-0.148513390	-0.10994594	0.13940561
citric_acid	-0.02197817	-0.14851339	-0.001495939	0.10661352	0.14156660
residual_sugar	0.27056607	-0.10994594	0.106613516	-0.16804783	0.21641566
chlorides	0.08308016	0.13940561	0.141566597	0.21641566	-0.03699737
free_sulfur_dioxide	0.06854074	0.05976442	0.031551022	0.32595496	0.10499147
total_sulfur_dioxide	0.08482380	0.14715360	0.066371202	0.30947157	0.12990603
density	0.23681052	-0.05259942	0.125176182	0.19325736	0.15330558
pH	-0.16734793	-0.11421838	-0.100175421	-0.01658089	-0.07880830
sulphates	-0.09313384	-0.07453851	0.029420451	0.07586081	0.02731841
alcohol	-0.31862532	-0.08516039	-0.158927533	-0.46732895	0.07573603
quality	-0.10418857	-0.26334723	-0.032540895	-0.24009150	0.01472680
	free_sulfur_dioxide	total_sulfur_dioxide	density	pH	sulphates
fixed_acidity	0.06854074	0.08482380	0.23681052	-0.16734793	-0.09313384
volatile_acidity	0.05976442	0.14715360	-0.05259942	-0.11421838	-0.07453851
citric_acid	0.03155102	0.06637120	0.12517618	-0.10017542	0.02942045
residual_sugar	0.32595496	0.30947157	0.19325736	-0.01658089	0.07586081
chlorides	0.10499147	0.12990603	0.15330558	-0.07880830	0.02731841
free_sulfur_dioxide	-0.13154824	0.11892696	0.43223944	0.09845151	0.08316064
total_sulfur_dioxide	0.11892696	0.22123322	0.39271383	-0.01033842	0.09181832
density	0.43223944	0.39271383	0.37097820	-0.03106763	0.07623783
pH	0.09845151	-0.01033842	-0.03106763	-0.02918059	-0.09761100
sulphates	0.08316064	0.09181832	0.07623783	-0.09761100	-0.03417243
alcohol	-0.29479856	-0.29661414	-0.37131749	-0.09529845	-0.14903536
quality	-0.14933013	-0.15569858	-0.14678516	0.04591611	0.02991532
	alcohol	quality			
fixed_acidity	-0.31862532	-0.10418857			
volatile_acidity	-0.08516039	-0.26334723			
citric_acid	-0.15892753	-0.03254090			
residual_sugar	-0.46732895	-0.24009150			
chlorides	0.07573603	0.01472680			
free_sulfur_dioxide	-0.29479856	-0.14933013			
total_sulfur_dioxide	-0.29661414	-0.15569858			
density	-0.37131749	-0.14678516			
pH	-0.09529845	0.04591611			
sulphates	-0.14903536	0.02991532			
alcohol	-0.17300156	-0.12029353			
quality	-0.12029353	-0.08504302			

اختلاف نسبتاً کمی دارند و میتوان مفید بودن مدل را پذیرفت.

یافتن گروه های پنهان در داده ها با تحلیل عاملی

بار های عاملی با روش درست نمایی:

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5
fixed_acidity	0.103			1.026	-0.138
volatile_acidity	0.131	0.172	-0.141		0.101
citric_acid				0.303	
residual_sugar		1.064		-0.116	-0.128
chlorides	-0.410	-0.104			
free_sulfur_dioxide			1.055		
total_sulfur_dioxide	-0.170	0.118	0.497		0.138
density	-0.411	0.648		0.130	0.116
pH	0.143	-0.105		-0.183	0.717
sulphates				0.112	0.361
alcohol	1.070			0.102	0.117
quality	0.524	0.111	0.126		

عامل یا فاکتور اول نشان دهنده ی بار های زیاد کلریدهای هوا، چگالی هوا، الکل در هوا و کیفیت هوا است. میتوان گفت که فاکتور اول نقش کلرید ها و چگالی و الکل هوا در کیفیت هوا را بررسی میکند.

فاکتور دوم نشان دهنده ی بار های زیاد شکر باقی مانده در هوا و چگالی هوا است. بنظر میرسد که رابطه ی این دو را بررسی میکند.

فاکتور سوم میزان دی اکسید سولفور آزاد و کل را بررسی میکند.

فاکتور چهارم نقش اسید ثابت در هوا را مورد بررسی قرار می دهد.

و فاکتور پنجم میزان pH را بررسی میکند.

از تابع fa هم برای تحلیل با روش مولفه های اصلی استفاده میکنیم:

	PA3	PA1	PA4	PA2	PA5	h2	u2	com
fixed_acidity	0.00	0.10	-0.05	0.85	0.02	0.727	0.2731	1.0
volatile_acidity	0.02	0.09	-0.07	-0.06	0.69	0.500	0.5002	1.1
citric_acid	0.03	0.03	0.12	0.36	-0.17	0.170	0.8300	1.7
residual_sugar	1.14	0.14	0.15	0.11	0.06	1.363	-0.3632	1.1
chlorides	0.02	0.37	0.09	0.05	0.04	0.151	0.8494	1.2
free_sulfur_dioxide	0.18	0.09	0.59	0.00	-0.08	0.391	0.6090	1.3
total_sulfur_dioxide	0.17	0.29	0.96	0.12	0.20	1.097	-0.0974	1.4
density	0.60	0.59	0.26	0.20	-0.03	0.822	0.1779	2.6
pH	-0.10	-0.08	0.11	-0.48	-0.07	0.269	0.7311	1.3
sulphates	-0.03	0.00	0.17	-0.05	-0.07	0.036	0.9637	1.6
alcohol	-0.27	-0.93	-0.17	-0.05	0.19	1.010	-0.0096	1.3
quality	-0.01	-0.52	0.04	-0.08	-0.22	0.327	0.6731	1.4

عامل سوم (PA3) نشان دهنده ی رابطه بین شکر موجود در هوا و چگالی هوا است و نقش الکل هوا را هم میتوان اضافه کرد به آن.

در عامل اول کلرید هوا، چگالی هوا، الکل هوا و کیفیت هوا بار های زیادی دارند و دی اکسید سولفور هم میتوان اضافه کرد.

این عامل را میتوان عامل تعیین کننده ی کیفیت هوا دانست.

عامل چهارم رابطی دی اکسید سولفور آزاد و کل را بررسی میکند.

در عامل دوم اسید ثابت بار زیادی دارد و با توجه به بار های عاملی میتوان گفت این عامل نقش اسید ثابت در هوا را مورد بررسی قرار می دهد.

و عامل پنجم هم اسیدیته فرار را بررسی میکند.

رده بندی کیفیت هوا: تحلیل ممیزی خطی و درجه دوم

همانطور که مشاهده کردید کیفیت هوا به ۷ نمره تقسیم شده است.

خلاصه ای از کیفیت هوا را مشاهده میکنید:

3	4	5	6	7	8	9	کیفیت هوا
20	163	1457	2198	880	175	5	تعداد روزها

حال برای انجام رده بندی از قاعده تست و ترین استفاده میکنیم و برای این کار از این کد ها استفاده میکنیم:

```
set.seed(1234)
```

```
training_sample=sample(c(TRUE,FALSE),nrow(x),replace = TRUE,prob = c(0.6,0.4))
```

```
train=x[training_sample, ]
```

```
test=x[!training_sample, ]
```

در اینجا ما با ۶۰ درصد از داده ها رده بندی انجام میدهیم.

اکنون با استفاده از این تابع `x.lm=lda(quality~.,data = train)` احتمال های پیشین و قاعده ی خطی و μ ها را بدست می آوریم.

تابع `lda` قاعده ی خطی با فرض مساوی بودن ماتریس کواریانس ها میبازد.

احتمال های پیشین برابر است با:

Prior probabilities of groups:

3	4	5	6	7	8	9
0.004083299	0.033278889	0.297468354	0.448754594	0.179665169	0.035728869	0.001020825

همانطور که ملاحظه میشود ۵ و ۶ و ۷ بیشترین احتمالات را دارند.

مقدار μ ها یا میانگین رده ها عبارت است از:

Group means:

	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide
3	7.600000	0.3332500	0.3360000	6.392500	0.05430000	53.32500
4	7.129448	0.3812270	0.3042331	4.628221	0.05009816	23.35890
5	6.933974	0.3020110	0.3376527	7.334969	0.05154633	36.43205
6	6.837671	0.2605641	0.3380255	6.441606	0.04521747	35.65059
7	6.734716	0.2627670	0.3256250	5.186477	0.03819091	34.12557
8	6.657143	0.2774000	0.3265143	5.671429	0.03831429	36.72000
9	7.420000	0.2980000	0.3860000	4.120000	0.02740000	33.40000
	total_sulfur_dioxide	density	pH	sulphates	alcohol	
3	170.6000	0.9948840	3.187500	0.4745000	10.34500	
4	125.2791	0.9942767	3.182883	0.4761350	10.15245	
5	150.9046	0.9952626	3.168833	0.4822032	9.80884	
6	137.0473	0.9939613	3.188599	0.4911056	10.57537	
7	125.1148	0.9924524	3.213898	0.5031023	11.36794	
8	126.1657	0.9922359	3.218686	0.4862286	11.63600	
9	116.0000	0.9914600	3.308000	0.4660000	12.18000	

و قاعده ی خطی آن ها به صورت زیر است:

Coefficients of linear discriminants:

	LD1	LD2	LD3	LD4	LD5	LD6
fixed_acidity	1.864850e-01	-2.767924e-01	-0.891126996	-0.24490813	-1.272544954	-0.332628327
volatile_acidity	-4.755577e+00	-7.217154e+00	-4.422485889	3.98626697	0.800959655	-0.325390191
citric_acid	-7.046494e-02	9.532036e-01	0.037884807	1.52679474	-1.353491004	-4.538797057
residual_sugar	1.890960e-01	1.120315e-01	-0.204335132	0.38027247	-0.301552196	-0.165564291
chlorides	-5.294565e-01	3.607816e+00	-7.720351920	-6.70585890	15.867529727	-2.668574152
free_sulfur_dioxide	1.060828e-02	4.827739e-03	-0.041649800	-0.02672595	0.018780352	0.014209319
total_sulfur_dioxide	-1.229068e-03	8.741603e-03	-0.005304268	0.01005809	-0.006497657	0.005690075
density	-3.445254e+02	-1.625747e+02	547.538798274	-922.13938757	990.798874200	159.781336415
pH	1.698011e+00	-9.577503e-01	-3.391073937	1.32542128	-5.527903366	-4.122631442
sulphates	1.610113e+00	-9.398279e-02	0.867088233	1.75433955	-3.698883397	5.119687195
alcohol	5.360439e-01	-2.897264e-01	0.385518802	-1.05758219	1.460492500	0.140246916

در اینجا ضریب d ها را داریم که ۶ تا ld کافی است.

از روی ۱ ld ها میتوان گفت که هرچه مقدار fix acidity و residual_sugar و free_sulfur_dioxide و pH و sulphates و alcohol بیشتر شود یعنی دارد به نفع رده اول رای می دهد و آن هایی که ضرایب منفی دارند اگر بیشتر شوند یعنی به نفع رده های دیگر رای میدهند.

خلاصه ی رده بندی با نمونه train به صورت زیر است:

	3	4	5	6	7	8	9
3	1	0	0	0	0	0	0
4	0	25	25	13	0	2	0
5	3	33	448	229	19	1	0
6	7	43	389	932	331	53	0
7	1	2	8	145	185	54	4
8	0	0	0	1	0	0	0
9	0	0	0	0	0	0	0

در جدول بالا تعداد داده هایی که به درستی رده بندی شده اند با رنگ سبز و تعداد داده هایی که اشتباه رده بندی شده اند با رنگ قرمز مشخص شده است.

همانطور که میبینیم فقط یکی از آن ها که واقعا جزو رده اول (۳) بودند به درستی پیش بینی شده اند که گفته بودیم جزو رده اول هستند. ۲۵ تای آن ها به درستی در رده دوم (۴) رده بندی شده است. ۳۳ تا از داده های ما جزو رده سوم (۵) بود و به اشتباه در رده دوم رده بندی کردیم و ۴۳ تای آن مربوط به رده ی چهارم (۶) است که در رده دوم قرار دارد.

در بین تمام رده ها داده های رده چهارم (۶) از بقیه رده ها بهتر رده بندی شده است.

خلاصه ی رده بندی با نمونه test به صورت زیر است:

	3	4	5	6	7	8	9
3	1	0	0	1	0	0	0
4	2	11	19	6	0	0	0
5	3	27	274	145	16	1	0
6	1	21	288	608	213	39	1
7	0	1	5	116	116	25	0
8	1	0	1	0	0	0	0
9	0	0	0	2	0	0	0

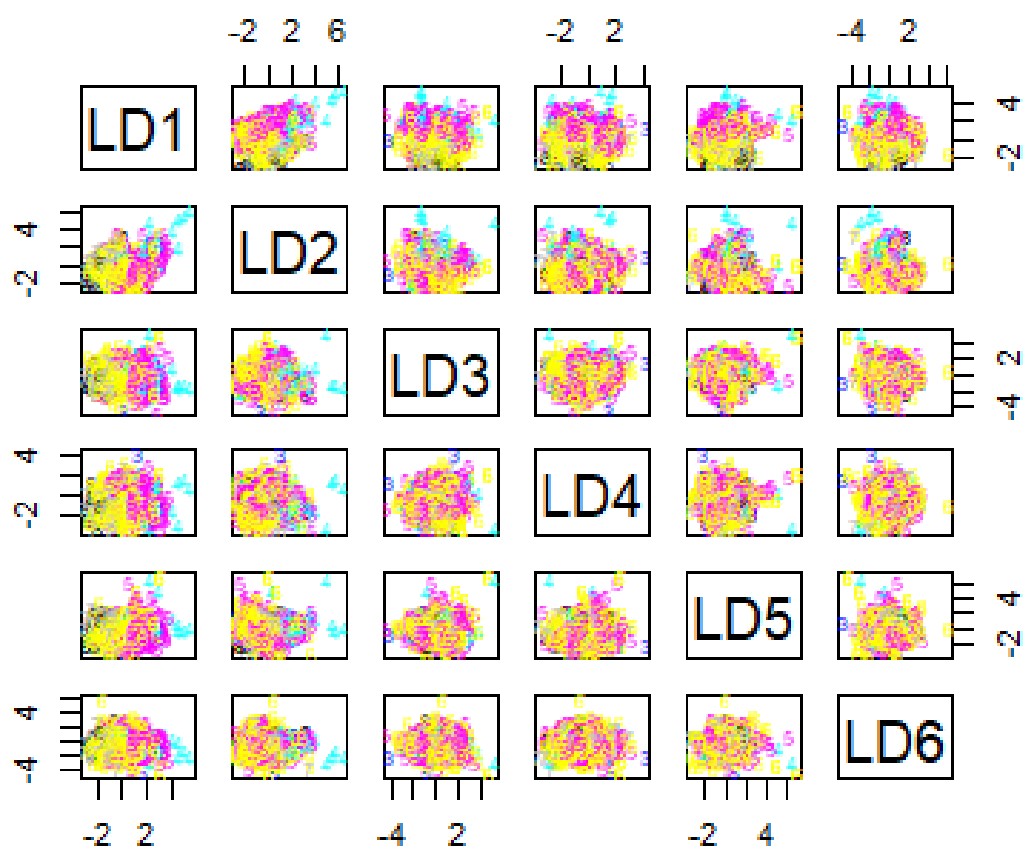
نرخ درستی رده بندی برابر است با:

3 4 5 6 7 8 9
0.5000000 0.2894737 0.5879828 0.5192143 0.4410646 0.0000000 0.0000000

یعنی رده اول ۵۰ درصد درست رده بندی شده است و رده دوم ۲۸ درصد و ...

و میزان خطای AER ما بر اساس رده بندی خطی (lda) برابر است با:

`appr_AER=1-sum(diag(table(test$lda,test$quality)))/nrow(test)= 0.480452`



با توجه به نمودار بالا فقط دو رده خوب از هم جدا می شوند و بقیه رده ها همپوشانی زیاد و جدایی کمی دارند.

کدهای استفاده شده

```
x=read.table("C:/Users/AH3134/Desktop/uni/PEYVASTE2/wind-quality-white1.txt",header = TRUE)
cormatrix=cor(x)
eig.pca <- eigen(cormatrix)
print(eig.pca$values)

plot(1:12,eig.pca$values,type="b")
abline(a=1,b=0)

fit=princomp(covmat=cormatrix)
fit1=princomp(x)
summary(fit)
fit$loadings
fit1$scores
fit2=prcomp(data=center=TRUE,scale. = TRUE)
summary(fit2)
screeplot(fit2,type="l")
p=cumsum(fit2$sdev^2/sum(fit2$sdev^2))
plot(p,xlab="pc",ylab="darsad variance ")
abline(h=0.98,col="blue")
abline(v=5,col="blue")

m=factanal(covmat = cormatrix,factors = 5,cor=TRUE,rotation = "varimax")
m=factanal(covmat = cormatrix,factors = 5,cor=TRUE,rotation ="promax")
l=m$loadings
psi=diag(m$uniquenesses)
rhat=l%*%t(l)+psi
r=m$correlation
library(psych)
```

```

fa(r=x,nfactors=5,rotate="none",fm="pa")

fa(r=x,nfactors=5,rotate="varimax",fm="pa")
#####

library(MASS)
library(klaR)
head(x)
summary(x)
table(x$quality)
set.seed(1234) ; training_sample=sample(c(TRUE,FALSE),nrow(x),replace = TRUE,prob = c(0.6,0.4))

train=x[training_sample, ]
test=x[!training_sample, ]
lda(quality~.,data = x)
x.lda=lda(quality~.,data = train)
plot(x.lda,col=as.integer(train$quality)+1)
#####

lda.train=predict(x.lda)
train$lda=lda.train$class
table(train$lda,train$quality)
#####
lda.test=predict(x.lda,test)
test$lda=lda.test$class
table(test$lda,test$quality)
#####
ct=table(test$lda,test$quality)
diag(prop.table(ct,1))
appr_AER=1-sum(diag(table(test$lda,test$quality)))/nrow(test)

```