

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 31.12.2020 13:36:44
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a481fda71c085

МИНОБРАЗОВАНИЯ И НАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра биомедицинской инженерии



УТВЕРЖДАЮ

Проректор по учебной работе
О.Г. Локтионова
« 31 » _____ 2017 г.

ОСНОВЫ АВТОМАТИЗИРОВАННОЙ ОБРАБОТКИ РЕЗУЛЬТАТОВ МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Методические указания к самостоятельной работе по дисциплинам
«Автоматизация обработки биомедицинской информации»,
«Стандартные программные средства в имитационном
моделировании биотехнических систем», «Прикладные пакеты
математической обработки данных» и «Автоматизация обработки
экспериментальных данных»

Курск 2017

УДК 004.93:61

Составители: О.В. Шаталова, К.Д.А. Кассим, С.А. Филист.

Рецензент

Доктор технических наук, профессор Р.А. Томакова

Основы автоматизированной обработки результатов медико-биологических исследований: методические указания к самостоятельной работе / Юго-Зап. гос. ун-т; сост.: О.В. Шаталова, К.Д.А. Кассим, С.А. Филист. Курск, 2017. 17 с.

Предназначено для студентов направления подготовки 12.03.04 «Биотехнические системы и технологии» и по специальности 30.05.03 «Медицинская кибернетика» по дисциплинам «Автоматизация обработки биомедицинской информации», «Стандартные программные средства в имитационном моделировании биотехнических систем», «Прикладные пакеты математической обработки данных» и «Автоматизация обработки экспериментальных данных». Может быть использована аспирантами, обучающимися по направленностям 05.11.13 – Системный анализ, управление и обработка информации и 05.11.17 – Приборы, системы и изделия медицинского назначения.

Текст печатается в авторской редакции

Подписано в печать 5.05.17. Формат 60×84 1/16. Бумага офсетная.

Усл. печ. л. 0,99. Уч.-изд. л. 0,89. Тираж 100 экз. Заказ 890.

Юго-Западный государственный университет.

305040, г. Курск, ул. 50 лет Октября, 94.

1 Самостоятельная работа студентов (СРС)

Исследование эффективности классификации двухальтернативной выборки геометрическими методами распознавания

В процессе выполнения работы необходимо сформировать две обучающих выборки. Каждая выборка представляет определенный класс индивидуумов, характеризующихся набором информативных признаков. Число объектов в обучающих выборках и число информативных признаков определяется индивидуальным заданием.

Затем проводится разведочный анализ, который заключается в исследовании гистограмм информативных признаков в многомерных пространствах и их сравнения (по классам).

Одномерные гистограммы по каждому признаку строятся для двух обучающих выборок, при этом шаг рассчитаем по формуле:

$$\text{Шаг} = (\max \text{ значение} - \min \text{ значение}) / (1 + \log_2 N),$$

где N – объем обучающей выборки.

В результате этого исследования исключаются те признаки, гистограммы которого имеют площадь перекрытия выше 80%.

Затем необходимо провести корреляционный анализ, устанавливающий линейную зависимость признаков, который должен подтвердить или не подтвердить верность исключения выбранного признака.

После этого исследуется три способа классификации данных, основанных на геометрических методах распознавания;

1 – построение разделяющей гиперплоскости;

2- определение расстояний Махаланобиса;

3- использование пакета дискриминантного анализа из Statistica.

В итоге формируются три решающих правила разделение совокупностей.

На следующем этапе формируются две контрольных выборки и проверяется эффективность решающих правил.

Обучающие и контрольные выборки формируются студентом самостоятельно на основании их статистических характеристик, приведенных в индивидуальном задании.

В заключении сравнивается эффективность решающих правил, разработанных на основе трех методов. Для наглядности сравнения решающих правил необходимо построить гистограммы, отражающие диагностическую чувствительность, диагностическую специфичность и диагностическую эффективность каждого решающего правила.

Индивидуальные задания

Таблица 1 - Исходные данные для первой совокупности

№ вар.	Нобуч/№контр.	Математические ожидания информативных признаков						Дисперсии информативных признаков					
		1	30/20	-18	112	101	495	10	51	100	500	100	1500
2	40/35	-21	109	140	491	17	49	81	145	39	1600	9	81
3	35/40	-21	100	132	513	20	-49	144	98	45	1460	9	144
4	15/30	20	95	120	499	16	-40	64	60	78	1000	18	64
5	18/25	26	73	66	509	17	-46	49	100	87	1000	20	49
6	25/19	-19	91	107	122	17	66	100	98	90	140	25	100
7	34/20	-15	68	93	122	23	-43	36	45	65	90	25	36
8	41/35	-19	72	78	127	13	-49	49	40	66	60	16	49
9	36/42	16	148	72	123	87	42	9	36	78	100	100	9
10	18/30	17	70	118	128	86	-60	16	26	76	98	100	16
11	19/25	-18	83	97	115	82	52	25	28	78	45	64	25
12	28/19	-21	107	141	124	80	-58	25	56	98	40	64	25
13	30/23	-22	112	97	127	70	55	49	80	100	36	49	49

Таблица 2 - Исходные данные для второй совокупности

№ вар.	Нобуч/№контр.	Математические ожидания информативных признаков						Дисперсии информативных признаков					
		1	20/40	-8	120	90	490	15	61	10	100	100	400
2	40/35	-25	209	150	441	27	69	8	145	39	500	9	49
3	42/40	-28	110	80	413	27	-29	14	98	45	300	9	36
4	19/30	30	85	128	409	19	-30	6	60	78	100	18	9
5	18/25	26	93	76	559	27	-56	4	100	87	100	20	49
6	35/19	-9	71	88	122	27	86	10	98	90	100	25	49
7	24/20	-5	70	89	132	33	-33	8	45	65	80	25	36
8	31/35	-9	92	60	147	23	-69	4	40	66	9	16	49
9	36/42	20	120	87	133	97	52	9	36	78	10	10	36
10	48/30	20	80	81	158	116	-50	16	26	76	9	10	16
11	19/25	-19	78	100	125	102	82	25	28	78	4	4	25
12	28/19	-29	120	131	134	89	-78	25	56	98	4	4	25
13	35/25	-28	140	87	147	80	59	4	80	100	6	9	49

Получение контрольных и обучающих выборок

Используя генератор случайных чисел пакета MathCad можно создать как обучающие, так и контрольные выборки с любым объемом и любой размерности пространства информативных признаков. Пример листа MathCad, в котором создается выборка из 28 объектов с нормально распределенным законом распределения признаков (признаковое пространство четырехмерное) показан на рисунке 1.

$$\begin{aligned}
 n & :- 28 & m & :- 4 & i & :- 0.. m - 1 & j & :- 0.. n - 1 \\
 s_0 & :- 225 & sd_0 & :- 16 & s_1 & :- 228 & sd_1 & :- 18 & s_2 & :- 231 \\
 sd_2 & :- 15 & s_3 & :- 235 & sd_3 & :- 16 \\
 \overrightarrow{V\lambda_i} & :- \text{ceil}(\text{rnorm}(n, s_i, sd_i))
 \end{aligned}$$

Рисунок 1 - Лист MathCad с генератором случайных чисел

Каждый информативный признак в выборке представлен вектором-столбцом $V\lambda$ из 28 элементов. Случайные числа, распределенные по нормальному закону распределения, генерируются функцией

$$\text{rnorm}(n, s, sd),$$

где n -число элементов в выборке, s – математическое ожидание информативного признака, sd -дисперсия информативного признака.

Варьируя эти параметры (s и sd) вы можете менять структуру распределения классов в признаковом пространстве и исследовать эффективность классификации при различных классовых структурах.

Функция ceil округляет случайное число до ближайшего целого.

Значок \longrightarrow обозначает операцию векторизации, то есть одновременное выполнение скалярной операции над всеми элементами вектора.

Пример полученной выборки в четырехмерном признаковом пространстве показан на рисунке 2.

$V\lambda_0 =$	<table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><th style="background-color: #cccccc;">0</th><td></td></tr> <tr><td>0</td><td>218</td></tr> <tr><td>1</td><td>215</td></tr> <tr><td>2</td><td>218</td></tr> <tr><td>3</td><td>210</td></tr> <tr><td>4</td><td>199</td></tr> <tr><td>5</td><td>226</td></tr> <tr><td>6</td><td>224</td></tr> <tr><td>7</td><td>234</td></tr> <tr><td>8</td><td>261</td></tr> <tr><td>9</td><td>238</td></tr> <tr><td>10</td><td>241</td></tr> <tr><td>11</td><td>239</td></tr> <tr><td>12</td><td>240</td></tr> <tr><td>13</td><td>236</td></tr> <tr><td>14</td><td>209</td></tr> <tr><td>15</td><td>227</td></tr> <tr><td>16</td><td>213</td></tr> <tr><td>17</td><td>237</td></tr> <tr><td>18</td><td>223</td></tr> <tr><td>19</td><td>215</td></tr> <tr><td>20</td><td>214</td></tr> <tr><td>21</td><td>217</td></tr> <tr><td>22</td><td>234</td></tr> <tr><td>23</td><td>222</td></tr> <tr><td>24</td><td>227</td></tr> <tr><td>25</td><td>246</td></tr> <tr><td>26</td><td>214</td></tr> <tr><td>27</td><td>226</td></tr> </table>	0		0	218	1	215	2	218	3	210	4	199	5	226	6	224	7	234	8	261	9	238	10	241	11	239	12	240	13	236	14	209	15	227	16	213	17	237	18	223	19	215	20	214	21	217	22	234	23	222	24	227	25	246	26	214	27	226	$V\lambda_1 =$	<table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><th style="background-color: #cccccc;">0</th><td></td></tr> <tr><td>0</td><td>248</td></tr> <tr><td>1</td><td>245</td></tr> <tr><td>2</td><td>176</td></tr> <tr><td>3</td><td>190</td></tr> <tr><td>4</td><td>232</td></tr> <tr><td>5</td><td>217</td></tr> <tr><td>6</td><td>207</td></tr> <tr><td>7</td><td>230</td></tr> <tr><td>8</td><td>242</td></tr> <tr><td>9</td><td>234</td></tr> <tr><td>10</td><td>229</td></tr> <tr><td>11</td><td>215</td></tr> <tr><td>12</td><td>221</td></tr> <tr><td>13</td><td>216</td></tr> <tr><td>14</td><td>226</td></tr> <tr><td>15</td><td>247</td></tr> <tr><td>16</td><td>232</td></tr> <tr><td>17</td><td>238</td></tr> <tr><td>18</td><td>242</td></tr> <tr><td>19</td><td>228</td></tr> <tr><td>20</td><td>234</td></tr> <tr><td>21</td><td>231</td></tr> <tr><td>22</td><td>236</td></tr> <tr><td>23</td><td>220</td></tr> <tr><td>24</td><td>216</td></tr> <tr><td>25</td><td>283</td></tr> <tr><td>26</td><td>223</td></tr> <tr><td>27</td><td>205</td></tr> </table>	0		0	248	1	245	2	176	3	190	4	232	5	217	6	207	7	230	8	242	9	234	10	229	11	215	12	221	13	216	14	226	15	247	16	232	17	238	18	242	19	228	20	234	21	231	22	236	23	220	24	216	25	283	26	223	27	205	$V\lambda_2 =$	<table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><th style="background-color: #cccccc;">0</th><td></td></tr> <tr><td>0</td><td>243</td></tr> <tr><td>1</td><td>226</td></tr> <tr><td>2</td><td>213</td></tr> <tr><td>3</td><td>216</td></tr> <tr><td>4</td><td>231</td></tr> <tr><td>5</td><td>232</td></tr> <tr><td>6</td><td>235</td></tr> <tr><td>7</td><td>239</td></tr> <tr><td>8</td><td>235</td></tr> <tr><td>9</td><td>221</td></tr> <tr><td>10</td><td>207</td></tr> <tr><td>11</td><td>253</td></tr> <tr><td>12</td><td>246</td></tr> <tr><td>13</td><td>214</td></tr> <tr><td>14</td><td>234</td></tr> <tr><td>15</td><td>257</td></tr> <tr><td>16</td><td>227</td></tr> <tr><td>17</td><td>201</td></tr> <tr><td>18</td><td>210</td></tr> <tr><td>19</td><td>212</td></tr> <tr><td>20</td><td>224</td></tr> <tr><td>21</td><td>232</td></tr> <tr><td>22</td><td>223</td></tr> <tr><td>23</td><td>227</td></tr> <tr><td>24</td><td>240</td></tr> <tr><td>25</td><td>235</td></tr> <tr><td>26</td><td>238</td></tr> <tr><td>27</td><td>214</td></tr> </table>	0		0	243	1	226	2	213	3	216	4	231	5	232	6	235	7	239	8	235	9	221	10	207	11	253	12	246	13	214	14	234	15	257	16	227	17	201	18	210	19	212	20	224	21	232	22	223	23	227	24	240	25	235	26	238	27	214
0																																																																																																																																																																																			
0	218																																																																																																																																																																																		
1	215																																																																																																																																																																																		
2	218																																																																																																																																																																																		
3	210																																																																																																																																																																																		
4	199																																																																																																																																																																																		
5	226																																																																																																																																																																																		
6	224																																																																																																																																																																																		
7	234																																																																																																																																																																																		
8	261																																																																																																																																																																																		
9	238																																																																																																																																																																																		
10	241																																																																																																																																																																																		
11	239																																																																																																																																																																																		
12	240																																																																																																																																																																																		
13	236																																																																																																																																																																																		
14	209																																																																																																																																																																																		
15	227																																																																																																																																																																																		
16	213																																																																																																																																																																																		
17	237																																																																																																																																																																																		
18	223																																																																																																																																																																																		
19	215																																																																																																																																																																																		
20	214																																																																																																																																																																																		
21	217																																																																																																																																																																																		
22	234																																																																																																																																																																																		
23	222																																																																																																																																																																																		
24	227																																																																																																																																																																																		
25	246																																																																																																																																																																																		
26	214																																																																																																																																																																																		
27	226																																																																																																																																																																																		
0																																																																																																																																																																																			
0	248																																																																																																																																																																																		
1	245																																																																																																																																																																																		
2	176																																																																																																																																																																																		
3	190																																																																																																																																																																																		
4	232																																																																																																																																																																																		
5	217																																																																																																																																																																																		
6	207																																																																																																																																																																																		
7	230																																																																																																																																																																																		
8	242																																																																																																																																																																																		
9	234																																																																																																																																																																																		
10	229																																																																																																																																																																																		
11	215																																																																																																																																																																																		
12	221																																																																																																																																																																																		
13	216																																																																																																																																																																																		
14	226																																																																																																																																																																																		
15	247																																																																																																																																																																																		
16	232																																																																																																																																																																																		
17	238																																																																																																																																																																																		
18	242																																																																																																																																																																																		
19	228																																																																																																																																																																																		
20	234																																																																																																																																																																																		
21	231																																																																																																																																																																																		
22	236																																																																																																																																																																																		
23	220																																																																																																																																																																																		
24	216																																																																																																																																																																																		
25	283																																																																																																																																																																																		
26	223																																																																																																																																																																																		
27	205																																																																																																																																																																																		
0																																																																																																																																																																																			
0	243																																																																																																																																																																																		
1	226																																																																																																																																																																																		
2	213																																																																																																																																																																																		
3	216																																																																																																																																																																																		
4	231																																																																																																																																																																																		
5	232																																																																																																																																																																																		
6	235																																																																																																																																																																																		
7	239																																																																																																																																																																																		
8	235																																																																																																																																																																																		
9	221																																																																																																																																																																																		
10	207																																																																																																																																																																																		
11	253																																																																																																																																																																																		
12	246																																																																																																																																																																																		
13	214																																																																																																																																																																																		
14	234																																																																																																																																																																																		
15	257																																																																																																																																																																																		
16	227																																																																																																																																																																																		
17	201																																																																																																																																																																																		
18	210																																																																																																																																																																																		
19	212																																																																																																																																																																																		
20	224																																																																																																																																																																																		
21	232																																																																																																																																																																																		
22	223																																																																																																																																																																																		
23	227																																																																																																																																																																																		
24	240																																																																																																																																																																																		
25	235																																																																																																																																																																																		
26	238																																																																																																																																																																																		
27	214																																																																																																																																																																																		
	$V\lambda_3 =$	<table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><th style="background-color: #cccccc;">0</th><td></td></tr> <tr><td>0</td><td>218</td></tr> <tr><td>1</td><td>218</td></tr> <tr><td>2</td><td>229</td></tr> <tr><td>3</td><td>214</td></tr> <tr><td>4</td><td>190</td></tr> <tr><td>5</td><td>256</td></tr> <tr><td>6</td><td>248</td></tr> <tr><td>7</td><td>238</td></tr> <tr><td>8</td><td>219</td></tr> <tr><td>9</td><td>221</td></tr> <tr><td>10</td><td>221</td></tr> <tr><td>11</td><td>240</td></tr> <tr><td>12</td><td>229</td></tr> <tr><td>13</td><td>256</td></tr> <tr><td>14</td><td>213</td></tr> <tr><td>15</td><td>242</td></tr> <tr><td>16</td><td>243</td></tr> <tr><td>17</td><td>264</td></tr> <tr><td>18</td><td>225</td></tr> <tr><td>19</td><td>244</td></tr> <tr><td>20</td><td>260</td></tr> <tr><td>21</td><td>206</td></tr> <tr><td>22</td><td>256</td></tr> <tr><td>23</td><td>219</td></tr> <tr><td>24</td><td>243</td></tr> <tr><td>25</td><td>219</td></tr> <tr><td>26</td><td>216</td></tr> <tr><td>27</td><td>242</td></tr> </table>	0		0	218	1	218	2	229	3	214	4	190	5	256	6	248	7	238	8	219	9	221	10	221	11	240	12	229	13	256	14	213	15	242	16	243	17	264	18	225	19	244	20	260	21	206	22	256	23	219	24	243	25	219	26	216	27	242																																																																																																																							
0																																																																																																																																																																																			
0	218																																																																																																																																																																																		
1	218																																																																																																																																																																																		
2	229																																																																																																																																																																																		
3	214																																																																																																																																																																																		
4	190																																																																																																																																																																																		
5	256																																																																																																																																																																																		
6	248																																																																																																																																																																																		
7	238																																																																																																																																																																																		
8	219																																																																																																																																																																																		
9	221																																																																																																																																																																																		
10	221																																																																																																																																																																																		
11	240																																																																																																																																																																																		
12	229																																																																																																																																																																																		
13	256																																																																																																																																																																																		
14	213																																																																																																																																																																																		
15	242																																																																																																																																																																																		
16	243																																																																																																																																																																																		
17	264																																																																																																																																																																																		
18	225																																																																																																																																																																																		
19	244																																																																																																																																																																																		
20	260																																																																																																																																																																																		
21	206																																																																																																																																																																																		
22	256																																																																																																																																																																																		
23	219																																																																																																																																																																																		
24	243																																																																																																																																																																																		
25	219																																																																																																																																																																																		
26	216																																																																																																																																																																																		
27	242																																																																																																																																																																																		

Рисунок 2 - Пример выборки из 28 элементов в четырехмерном признаковом пространстве

Порядок построения линейной разделяющей гиперплоскости

По полученным обучающим выборкам найдем разделяющую гиперплоскость, которая проходит через середину отрезка, соединяющего центроиды двух обучающих выборок и перпендикулярна к нему.

Построим линейную разделяющую поверхность для двух классов.

Если вектор X_i характеризует i -й объект первого класса, а вектор Y_j характеризует j -й объект второго класса, то координаты

центроид A и B , а, следовательно, и отрезка AB , их соединяющего, определяются как

$$M[X] \text{ и } M[Y].$$

Координаты точки C , которая лежит на середине отрезка AB , определяются по формуле

$$(M[x_1] + M[y_1])/2; (M[x_2] + M[y_2])/2; \dots (M[x_N] + M[y_N])/2,$$

где N – число информативных признаков или размерность признакового пространства.

Если плоскость перпендикулярна вектору $\mathbf{n}(a; b; c)$, например, в трехмерном пространстве, то уравнение плоскости в этом пространстве записывается как

$$ax + by + cz + d = 0$$

Уравнение плоскости, перпендикулярной вектору $\mathbf{n}(a; b; c)$ и проходящей через точку $(x_0; y_0; z_0)$ записывается как

$$a(x - x_0) + b(y - y_0) + c(z - z_0) + d = 0.$$

Чтобы перейти от вектора AB к вектору \mathbf{n} , необходимо из координат $M[X_i]$ вычесть координаты $M[Y_i]$, то есть из координат точки A вычесть координаты точки B .

Дискриминантный анализ

Основная цель

Дискриминантный анализ используется для принятия решения о том, какие переменные различают (дискриминируют) две или более возникающие совокупности (группы) [1]. Например, некий исследователь в области образования может захотеть исследовать, какие переменные относят выпускника средней школы к одной из трех категорий: (1) поступающий в колледж, (2) поступающий в профессиональную школу или (3) отказывающийся от дальнейшего образования или профессиональной подготовки. Для этой цели исследователь может собрать данные о различных переменных, связанных с учащимися школы. После выпуска большинство учащихся, естественно, должно попасть в одну из названных категорий. Затем можно использовать *Дискриминантный анализ* для определения того, какие переменные дают наилучшее предсказание выбора учащимися дальнейшего пути.

Медик может регистрировать различные переменные, относящиеся к состоянию больного, чтобы выяснить, какие переменные лучше предсказывают, что пациент, вероятно, выздоровел полностью (группа 1), частично (группа 2) или совсем не выздоровел (группа 3). Биолог может записать различные характеристики сходных типов (групп) цветов, чтобы затем провести анализ дискриминантной функции, наилучшим образом разделяющей типы или группы.

Классификация

Другой главной целью применения дискриминантного анализа является проведение классификации. Как только модель установлена и получены дискриминирующие функции, возникает вопрос о том, как хорошо они могут *предсказывать*, к какой совокупности принадлежит конкретный образец?

Априорная и апостериорная классификация. Прежде чем приступить к изучению деталей различных процедур оценивания, важно уяснить, что эта разница ясна. Обычно, если вы оцениваете на основании некоторого множества данных дискриминирующую функцию, наилучшим образом разделяющую совокупности, и затем используете *те же самые* данные для оценивания того, какова точность вашей процедуры, то вы во многом полагаетесь на волю случая. В общем случае, получают, *конечно* худшую классификацию для образцов, не использованных для оценки дискриминантной функции. Другими словами, классификация действует лучшим образом для выборки, по которой была проведена оценка дискриминирующей функции (*апостериорная* классификация), чем для свежей выборки (*априорная* классификация). Трудности с (*априорной*) классификацией будущих образцов заключается в том, что никто не знает, что может случиться. Намного легче классифицировать уже имеющиеся образцы. Поэтому оценивание качества процедуры классификации никогда не производят по той же самой выборке, по которой была оценена дискриминирующая функция. Если желают использовать процедуру для классификации будущих образцов, то ее следует "испытать" (произвести кросс-проверку) на новых объектах.

Функции классификации. Функции классификации не следует путать с дискриминирующими функциями. Функции классификации предназначены для определения того, к какой

группе наиболее вероятно может быть отнесен каждый объект. Имеется столько же функций классификации, сколько групп. Каждая функция позволяет вам для каждого образца и для каждой совокупности вычислить *веса классификации* по формуле:

$$S_i = c_i + w_{i1} * x_1 + w_{i2} * x_2 + \dots + w_{im} * x_m$$

В этой формуле индекс i обозначает соответствующую совокупность, а индексы $1, 2, \dots, m$ обозначают m переменных; c_i являются константами для i -ой совокупности, w_{ij} - веса для j -ой переменной при вычислении показателя классификации для i -ой совокупности; x_j - наблюдаемое значение для соответствующего образца j -ой переменной. Величина S_i является результатом показателя классификации.

Поэтому вы можете использовать функции классификации для прямого вычисления показателя классификации для некоторых новых значений.

Классификация наблюдений. Как только вы вычислили показатели классификации для наблюдений, легко решить, как производить классификацию наблюдений. В общем случае наблюдение считается принадлежащим той совокупности, для которой получен наивысший показатель классификации (кроме случая, когда вероятности *априорной* классификации становятся слишком малыми; см. ниже). Поэтому, если вы изучаете выбор карьеры или образования учащимися средней школы после выпуска (поступление в колледж, в профессиональную школу или получение работы) на основе нескольких переменных, полученных за год до выпуска, то можете использовать функции классификации, чтобы предсказать, что наиболее вероятно будет делать каждый учащийся после выпуска. Однако вы хотели бы определить *вероятность*, с которой учащийся сделает предсказанный выбор. Эти вероятности называются *апостериорными*, и их также можно вычислить. Однако для понимания, как эти вероятности вычисляются, вначале рассмотрим так называемое *расстояние Махаланобиса*.

Расстояние Махаланобиса. В общем, расстояние Махаланобиса является мерой расстояния между двумя точками в пространстве, определяемым двумя или более *коррелированными* переменными. Например, если имеются всего две некоррелированных переменные, то вы можете нанести точки (образцы) на стандартную 2М диаграмму рассеяния. Расстояние Махаланобиса между точками будет в этом

случае равно расстоянию Евклида, т.е. расстоянию, измеренному, например, рулеткой. Если имеются три некоррелированные переменные, то для определения расстояния вы можете по-прежнему использовать рулетку (на 3М диаграмме). При наличии более трех переменных вы не можете более представить расстояние на диаграмме. Также и в случае, когда переменные коррелированы, то оси на графике могут рассматриваться как *неортогональные* (они уже не направлены под прямыми углами друг к другу). В этом случае простое определение расстояния Евклида не подходит, в то время как расстояние Махаланобиса является адекватно определенным в случае наличия корреляций.

Расстояние Махаланобиса и классификация. Для каждой совокупности в выборке вы можете определить положение точки, представляющей средние для всех переменных в многомерном пространстве, определенном переменными рассматриваемой модели. Эти точки называются *центроидами* группы. Для каждого наблюдения вы можете затем вычислить его расстояние Махаланобиса от каждого центроида группы. Снова, вы признаете наблюдение принадлежащим к той группе, к которой он ближе, т.е. когда расстояние Махаланобиса до нее минимально.

Апостериорные вероятности классификации. Используя для классификации расстояние Махаланобиса, вы можете теперь получить вероятность того, что образец принадлежит к конкретной совокупности. Это значение будет не вполне точным, так как распределение вокруг среднего для каждой совокупности будет не в точности нормальным. Так как принадлежность каждого образца вычисляется по априорному знанию модельных переменных, эти вероятности называются *апостериорными* вероятностями. Короче, *апостериорные* вероятности - это вероятности, вычисленные с использованием знания значений других переменных для образцов из частной совокупности. Некоторые пакеты автоматически вычисляют эти вероятности для всех наблюдений (или для выбранных наблюдений при проведении кросс-проверки).

Априорные вероятности классификации. Имеется одно дополнительное обстоятельство, которое следует рассмотреть при классификации образцов. Иногда вы знаете заранее, что в одной из групп имеется больше наблюдений, чем в другой. Поэтому *априорные* вероятности того, что образец принадлежит такой группе, выше. Например, если вы знаете заранее, что 60%

выпускников вашей средней школы обычно идут в колледж, (20% идут в профессиональные школы и остальные 20% идут работать), то вы можете уточнить предсказание таким образом: при всех других равных условиях более вероятно, что учащийся поступит в колледж, чем сделает два других выбора. Вы можете установить различные *априорные* вероятности, которые будут затем использоваться для уточнения результатов классификации наблюдений (и для вычисления *апостериорных* вероятностей).

На практике, исследователю необходимо задать себе вопрос, является ли неодинаковое число наблюдений в различных совокупностях в первоначальной выборке отражением истинного распределения в популяции, или это только (случайный) результат процедуры выбора. В первом случае вы должны положить *априорные* вероятности пропорциональными объемам совокупностей в выборке; во втором - положить *априорные* вероятности одинаковыми для каждой совокупности. Спецификация различных *априорных* вероятностей может сильно влиять на точность классификации.

Итог классификации. Общим результатом, на который следует обратить внимание при оценке качества текущей функции классификации, является *матрица классификации*. Матрица классификации содержит число образцов, корректно классифицированных (на диагонали матрицы) и тех, которые попали не в свои совокупности (группы).

Другие предостережения. При повторной итерации *апостериорная* классификация того, что случилось в прошлом, не очень трудна. Нетрудно получить очень хорошую классификацию тех образцов, по которым была оценена функция классификации. Для получения сведений, насколько хорошо работает процедура классификации на самом деле, следует классифицировать (*априорно*) *различные* наблюдения, то есть, наблюдения, которые не использовались при оценке функции классификации. Вы можете гибко использовать условия отбора для включения или исключения из вычисления наблюдений, поэтому матрица классификации может быть вычислена по "старым" образцам столь же успешно, как и по "новым". Только классификация новых наблюдений позволяет определить качество функции классификации (см. также кросс-проверку); классификация старых наблюдений позволяет

лишь провести успешную диагностику наличия выбросов или области, где функция классификации кажется менее адекватной.

Итог. В общем, *Дискриминантный анализ* - это очень полезный инструмент для поиска переменных, позволяющих относить наблюдаемые объекты в одну или несколько реально наблюдаемых групп, (2) - для классификации наблюдений в различные группы.

Оценка эффективности методов распознавания

В качестве расчетных показателей качества диагностических решающих правил используется: диагностическая чувствительность (ДЧ), диагностическая специфичность (ДС), прогностическая значимость положительных результатов (ПЗ⁺), прогностическая значимость отрицательных результатов (ПЗ⁻), диагностическая эффективность решающего правила (ДЭ) [2].

Эти показатели вычислялись по данным таблицы распределений результатов контрольных испытаний (таблица 3).

Таблица 3 - Таблица контрольных испытаний

Обследуемые	Результаты срабатывания правил		Всего
	положительные	отрицательные	
n_{ω_r}	ИП	ЛО	ИП+ЛО
n_{ω_0}	ЛП	ИО	ЛП+ИО
Всего	ИП+ЛП	ЛО+ИО	ИП+ЛП+ЛО+ИО

Примечание: r – номер класса исследуемого заболевания; n_{ω_r} - количество людей в контрольной выборке в исследуемом классе заболеваний; n_{ω_0} - количество здоровых людей в контрольной выборке; ИП – истинно положительный результат равный количеству людей класса ω_r правильно классифицируемых рассматриваемым правилом; ЛП – ложно положительный результат равный количеству людей класса ω_0 ошибочно отнесенных решающим правилом к классу ω_r ; ЛО – ложно отрицательный результат: количество людей класса ω_r ошибочно отнесенных решающим правилом к классу ω_0 ; ИО – истинно отрицательный

результат: количество людей класса ω_0 правильно классифицируемых решающим правилом.

Для приведенных в таблице 3 обозначений расчет показателей качества осуществляется в соответствии с выражениями:

$$\left\{ \begin{array}{l} \text{ДЧ} = \text{ИП} / n_{\omega_r} \\ \text{ДС} = \text{ИО} / n_{\omega_0} \\ \text{ПЗ}^+ = \text{ИП} / (\text{ИП} + \text{ЛП}) \\ \text{ПЗ}^- = \text{ИО} / (\text{ЛО} + \text{ИО}) \\ \text{ДЭ} = (\text{ИП} + \text{ИО}) / (\text{ИП} + \text{ЛП} + \text{ЛО} + \text{ИО}) \end{array} \right. \quad (1)$$

2 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

2.1 Основная учебная литература

1. Барботько, Анатолий Иванович. Статистические алгоритмы обработки результатов экспериментальных исследований в машиностроении [Текст] : [учебное пособие для студентов высших учебных заведений по направлению "Конструкторско-технологическое обеспечение машиностроительных производств"] / А. И. Барботько. - Старый Оскол : ТНТ, 2014. - 404 с. : табл. ; 21 см. - Библиогр.: с. 332-335. - 1000 экз. - ISBN 978-5-94178-452-3 (в пер.) : 580.00 р.

2. Статистика [Текст] : учебник для бакалавров / В. С. Мхитарян [и др.] ; под ред. проф. В. С. Мхитаряна. - Москва : Юрайт, 2015. - 590 с. : ил. - (Бакалавр. Базовый курс). - Библиогр.: с. 589-590 (20 назв.). - ISBN 978-5-9916-2411-4 (в пер.) : 744.67 р.

2.2 Дополнительная учебная литература

3. Апальков, Владимир Васильевич. Основы моделирования цифровой обработки сигналов в среде MATLAB [Текст] : учебное пособие / В. В. Апальков, Р. А. Томакова, Н. Н. Елишев ; Юго-Зап. гос. ун-т. - Курск : ЮЗГУ, 2015. - 136 с. + 1 эл. опт. диск (CD-ROM). - Библиогр.: с. 135. - 1000 экз. - ISBN 978-5-905979-66-8 : 350.00 р.

4. Комарова, Е. С. Парный регрессионный анализ [Электронный ресурс] : учебное пособие / Е. С. Комарова. - М. ; Берлин : Директ-Медиа, 2015. - 59 с. : ил. - ISBN 978-5-4475-4493-5 : Б. ц. // Режим доступа - <http://biblioclub.ru/>

5. Потехин, Д. С. Разработка систем цифровой обработки сигналов на базе ПЛИС [Текст] / Д. С. Потехин, И. Е. Тарасов. - М. : Горячая линия - Телеком, 2007. - 248 с. : ил. - (Современная электроника). - ISBN 978-5-93517-3 41-7 : 110.00 р.

6. Рангайян, Р. М. Анализ биомедицинских сигналов. Практический подход [Текст] : учебное пособие / Р. М. Рангайян. - М. : Физматлит, 2007. - 440 с. - ISBN 978-5-9221-07 30-3 : 200.00 р.

7. Сальников, Игорь Иванович. Растровые пространственно-временные сигналы в системах анализа изображений [Текст] / И. И. Сальников. - М. : Физматлит, 2009. - 248 с. - ISBN 978-5-9221-11 26-3 : 240.00 р.

8. Сергиенко, А. Б. Цифровая обработка сигналов [Текст] : учебное пособие / А. Б. Сергиенко. - 2-е изд. - СПб. : Питер, 2006. - 751 с. - (Учебник для вузов). - ISBN 5-469-00816-9 : 253.00 р.

9. Чобану, Михаил Константинович. Многомерные многоскоростные системы обработки сигналов [Текст] : [монография] / М. К. Чобану. - М. : Техносфера, 2009. - 480 с. - (Мир цифровой обработки). - ISBN 978-5-94836-2 33-5 : 235.00 р.

10. Яне, Бернд. Цифровая обработка изображений [Комплект] : [учебное пособие] / пер. с англ. А. М. Измайловой. - М. : Техносфера, 2007. - 584 с. : ил. + 1 эл. опт. диск (CD-ROM). - (Мир цифровой обработки. XI. 06). - Библиогр.: с. 575-583 (221 назв.). - ISBN 978-5-94836-1 22-2 : 285.00 р.

2.3 Перечень методических указаний

1. Основы автоматизированной обработки результатов медико-биологических исследований [Электронный ресурс] : методические указания к лабораторным работам по дисциплинам «Автоматизация обработки биомедицинской информации», «Стандартные программные средства в имитационном моделировании биотехнических систем» и «Автоматизация обработки экспериментальных данных» / Юго-Зап. гос. ун-т ; сост.: О. В. Шаталова, К. Д. А. Кассим, С. А. Филист. - Электрон. текстовые дан. (1 693 КБ). - Курск : ЮЗГУ, 2017. - 132 с. - Б. ц.

2. Основы автоматизированной обработки результатов медико-биологических исследований [Электронный ресурс] : методические указания к практическим занятиям по дисциплинам «Автоматизация обработки биомедицинской информации» и «Стандартные программные средства в имитационном

моделировании биотехнических систем» / Юго-Зап. гос. ун-т ; сост.: О. В. Шаталова, К. Д. А. Кассим, С. А. Филист. - Электрон. текстовые дан. (1 323 КБ). - Курск : ЮЗГУ, 2017. - 85 с. - Б. ц.

3. Основы автоматизированной обработки результатов медико-биологических исследований [Электронный ресурс] : методические указания к самостоятельной работе по дисциплинам «Автоматизация обработки биомедицинской информации», «Стандартные программные средства в имитационном моделировании биотехнических систем», «Прикладные пакеты математической обработки данных» и «Автоматизация обработки экспериментальных данных» / Юго-Зап. гос. ун-т ; сост.: О. В. Шаталова, К. Д. А. Кассим, С. А. Филист. - Электрон. текстовые дан. (437 КБ). - Курск : ЮЗГУ, 2017. - 17 с. - Б. ц.

2.4 Другие учебно–методические материалы

Отраслевые научно-технические журналы в библиотеке университета:

Биомедицинская радиоэлектроника

Биотехносфера

Медицинская техника

<https://www.youtube.com/watch?v=fxvxaTyZ6vw> – Обучающее видео «Проектирование цифровых фильтров в MATLAB»

<https://www.youtube.com/watch?v=reRuNxxhUuRs> – Обучающее видео «Вебинар "STATISTICA для медицинских приложений: актуальные задачи"»

https://www.youtube.com/watch?v=J_hGJ7wYCr4 – Обучающее видео «Интерактивное построение графиков в MATLAB»

3 Перечень ресурсов информационно–телекоммуникационной сети Интернет, необходимые для освоения дисциплины

1. www.statsoft.ru - STATSOFT

2. www.exponenta.ru/soft/Statist/Statist.asp - Образовательный математический сайт

3. <http://www.physionet.org/> - Физиологические сигналы

4. <http://www.lib.swsu.ru/> - Электронная библиотека ЮЗГУ

5. <http://www.biblioclub.ru> - Электронно-библиотечная система «Университетская библиотека online»

6. «"IPRbooks" <http://iprbookshop.ru> - Электронно-библиотечная система

7. <http://cyberleninka.ru> - Электронный портал

8. http://www.statsoft.ru/resources/statistica_text_book.php - Электронный учебник STATSOFT

4 Методические указания для обучающихся по освоению дисциплины

Основными видами аудиторной работы студента при изучении дисциплины *«Автоматизация обработки биомедицинской информации»* являются *лекции, лабораторные и практические занятия*. Студент не имеет права пропускать занятия без уважительных причин.

На лекциях излагаются и разъясняются основные понятия темы, связанные с ней теоретические и практические проблемы, даются рекомендации для самостоятельной работы. В ходе лекции студент должен внимательно слушать и конспектировать материал.

Изучение наиболее важных тем или разделов дисциплины завершают *лабораторные и практические занятия*, которые обеспечивают: контроль подготовленности студента; закрепление учебного материала; приобретение опыта устных публичных выступлений, ведения дискуссии, в том числе аргументации и защиты выдвигаемых положений и тезисов.

Практическому и лабораторному занятиям предшествует самостоятельная работа студента, связанная с освоением материала, полученного на лекциях, и материалов, изложенных в учебниках и учебных пособиях, а также литературе, рекомендованной преподавателем.

По согласованию с преподавателем или по его заданию студенты готовят рефераты по отдельным темам дисциплины, выступают на занятиях с докладами. Основу докладов составляет, как правило, содержание подготовленных студентами рефератов.

Качество учебной работы студентов преподаватель оценивает по результатам тестирования, собеседования, защиты отчетов по *практическим и лабораторным работам*, а также по результатам рубежных тестов.

Преподаватель уже на первых занятиях объясняет студентам, какие формы обучения следует использовать при самостоятельном

изучении дисциплины *«Автоматизация обработки биомедицинской информации»*: конспектирование учебной литературы и лекции, составление словарей понятий и терминов и т. п.

В процессе обучения преподаватели используют активные формы работы со студентами: чтение лекций, привлечение студентов к творческому процессу на лекциях, промежуточный контроль путем отработки студентами пропущенных лекций, участие в групповых и индивидуальных консультациях (собеседовании). Эти формы способствуют выработке у студентов умения работать с учебником и литературой. Изучение литературы составляет значительную часть самостоятельной работы студента. Это большой труд, требующий усилий и желания студента. В самом начале работы над книгой важно определить цель и направление этой работы. Прочитанное следует закрепить в памяти. Одним из приемов закрепления освоенного материала является конспектирование, без которого немыслима серьезная работа над литературой. Систематическое конспектирование помогает научиться правильно, кратко и четко излагать своими словами прочитанный материал.

Самостоятельную работу следует начинать с первых занятий. От занятия к занятию нужно регулярно прочитывать конспект лекций, знакомиться с соответствующими разделами учебника, читать и конспектировать литературу по каждой теме дисциплины. Самостоятельная работа дает студентам возможность равномерно распределить нагрузку, способствует более глубокому и качественному усвоению учебного материала. В случае необходимости студенты обращаются за консультацией к преподавателю по вопросам дисциплины *«Автоматизация обработки биомедицинской информации»* с целью усвоения и закрепления компетенций.

Основная цель самостоятельной работы студента при изучении дисциплины *«Автоматизация обработки биомедицинской информации»* - закрепить теоретические знания, полученные в процессе лекционных занятий, а также сформировать практические навыки самостоятельного анализа особенностей дисциплины.