مجلة القادسية لعلوم الحاسوب والرياضيات المجلد (3) العدد (1) السنة (2011)

الصفحة 40-50

المؤتمر العلمي الثاني
لكلية علوم الحاسوب
والرياضيات
16-18/اذار/2010

إقتراح طريقة LASS معدلّة كقيد جزاء إضافي لتقدير المركبات الرئيسة مع التطبيق

فراس أحمد محمد أحمد　　*عمر عبدالمحسن علي*　　*سعيد حميد لطيف*

قسم الأحصاء

جامعة بغداد

الخلاصة

يتناول هذا البحث طريقة تقليص المركبات الرئيسة مشابهة لتلك المستعملة في الأنحدار المتعدد وهي طريقة "تقليص وأختيار أقل فروق مطلقة: LASS". والهدف هنا هو تكوين تراكيب خطية غير مرتبطة لمجموعة جزئية من المتغيرات التوضيحية والتي قد تعاني من مشكلة التعدد الخطي بدلاً عن أخذها جميعاً، (أي جميع الـ K) لتلك المتغيرات. أذ سيقوم هذا التقليص الجديد بتصفير (أي جعلها تساوي الصفر) بعض المعاملات، بعد تطبيق بعض القيود عليها عن طريق معلمة التناوب t والتي توازن بين مقدار التحيز والتباين من جهة، وبحيث لاتتجاوز حداً مقبولاً من معيار نسبة المساهمة المئوية للتباين المفسر لتلك المركبات. وقد تم بيان ذلك عن طريق معيار متوسط مربعات الخطأ MSE في حالة الأنحدار، وبنسبة المساهمة المئوية للتباين المفسر في حالة المركبات الرئيسة.

1. المقدمة وهدف البحث

يعدّ تحليل المركبات الرئيسة (Principal Component Analysis: PCA)، من أدوات متعدد المتغيرات الشائعة في تحليل البيانات وإختزال أبعادها وعمل التدوير لها. فالمركبات الرئيسة تقوم بإيجاد تراكيب خطية لـ k من المتغيرات التوضيحية X's، مثلاً، وتقوم هذه التراكيب بدورها بالأستحواذ على أكبر تباين موجود في البيانات. ومن المعلوم أن هذا الأجراء سيتم إنجازه عن طريق تجزئة Decomposition القيم الذاتية Eigen Values λ_i لمصفوفة التباين والتباين المشترك S (أو لمصفوفة الأرتباطات البسيطة R)، ويتم ترتيب المركبات الرئيسة المستخرجة تنازلياً من الأكبر الى الأصغر.

إن أهمية المركبات الرئيسة ترجع الى عوامل عدة، منها: أولاً، من خلال السيطرة على إتجاهات التباين الأعظم للبيانات، إذ ستقدم المركبات الرئيسة طريقة لأختزال البيانات مع أقل فقدان ممكن للمعلومات. ثانياً، إن المركبات الرئيسة غير مرتبطة، وهذا يساعد كثيراً في عملية التفسير وبالتالي يساعد في التحليل الأحصائي بالأخص في حالة وجود مشكلة التعدد الخطي multicollinearity بين الـ X's. إلاّ إنه بالمقابل هناك مآخذ على المركبات الرئيسة ومنها: إن التركيز في المركبات الرئيسة يكون عادةً على تراكيب خطية من جميع المتغيرات التوضيحية وهذا يسبب بعض الصعوبة في تفسير النتائج[2]، بسبب عدم توضيح أي من هذه المركبات كانت مساهمتها أضعف (أو أقوى) في تفسير التباين الإجمالي.

لذا يهدف هذا البحث الى الوصول الى طريقة تقوم بعمل تراكيب خطية مختزلة من مجموعة جزئية من المتغيرات التوضيحية الأصلية وليس جميعها. وستقوم هذه التراكيب الجديدة بإختزال أكثر للوصول الى أبعاد من درجة أوطأ تقوم بتوضيح وتفسير معظم التباين الموجود في البيانات بعد وضع بعض القيود الأضافية على مسألة تقدير المركبات الرئيسة. ويتم عمل ذلك دون المساس بجوهر التحليل فيبقى مقدار التباين المفسّر ضمن مداه المقبول. ويتم ذلك بتطبيق على بيانات إدارية تتعلق بالقيادة الستراتيجية. وللوصول الى هذا الهدف تم تقسيم البحث الى أربعة أجزاء: يستهل في الجزء الأول، المقدمة وهدف البحث. أما في الجزء الثاني، فقد تم عرض مايخص الجانب النظري لطريقة Least Absolute Shrinkage and Selection: LASS مع أقتراح حالة معدلة لها. وفي الجزء الثالث، تم تقديم الجانب التطبيقي المستند الى بيانات إدارية تخص القيادة الستراتيجية. وفي الجزء الأخيراً، تم تلخيص الأستنتاجات التي أفرزها البحث ووضع قائمة بالمصادر.

2. الجانب النظري

لقد تم تقديم إمكانية إستعمال جزء من المركبات الرئيسة في التحليل، وكما موضح في الفقرات اللاحقة.

1.2 تحليل المركبات الرئيسة Principal Components Analysis: PCA

يقوم هذا التحليل على مبدأ تقليص عدد كبير من المتغيرات التوضيحية X's الى عدد أقل من المركبات النظرية غير المرتبطة والتي تدعى "المركبات الرئيسة" مع ضمان أقل خسارة ممكنة في المعلومات. وستكون المركبات الرئيسة عبارة عن تراكيب خطية من هذه الـ X's، ومن الأهمية بمكان أن تكون وحدات القياس متساوية، وفيما عدا ذلك يكون علينا إستعمال القيم المعيارية لتحويل مصفوفة X ذات الدرجة (n×k)، $X=[x_{ij}]$; $i=1, 2, \ldots, n$; $j=1, 2, \ldots, k$ ، الى مصفوفة القيم المعيارية Z ذات الدرجة (n×k) أيضاً، إذ أن: $Z=[z_{ij}]$ ، وإن[4],[3],[2] :

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{\sqrt{S_{ij}}} \qquad \ldots(1)$$

وإن: $S_{ij} = \sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)^2$ ، $\sum_{i=1}^{n} Z_{ij} = 0$ ، $\sum_{i=1}^{n} Z_{ij}^2 = 1$ ، $\sum_{j=1}^{K}\sum_{i=1}^{n} Z_{ij}^2 = K$

وبذلك فإن المصفوفة Z'Z ستكون عبارة عن مصفوفة الأرتباطات البسيطة R لما بين المتغيرات التوضيحية: $X_1, X_2, \ldots, X_K$ ، أي أن: R=Z'Z.

وسيتم إستخراج القيم الذاتية Eigen Values والتي يرمز لها بالرمز λ_j ، وكذلك المتجهات الذاتية Eigen Vectors والتي يرمز لها بالرمز P_j.

وأن: $a'_j=[a_{1j}\ a_{2j}\ a_{3j}\ \ldots\ a_{kj}]$. وتكون قيم λ_j عبارة عن الحلول للمعادلة[3] [5],[4],:

$$|R - \lambda I| = 0 \qquad \ldots(2)$$

وإن:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0$$

بحيث تقترن كل قيمة كامنة λ_j بمتجه كامن a_j بحيث يحقق مجموعة المعادلات[3],[5],[4]:

$$(R - \lambda_j I)\, a_j = 0 \qquad ; j = 1, 2, 3, \ldots, k \qquad \ldots(3)$$

وبشرط: $a'_j a_j=1$. وستستعمل المتجهات a_j بعد ذلك في إعادة التعبير عن المركبات الرئيسة F_j بدلالة المتغيرات القياسية Z_j ، وبالصيغة الآتية:

$$F_j = a_{1j}Z_1 + a_{2j}Z_2 + \ldots + a_{kj}Z_k = a'_j\underline{Z} \qquad \ldots(4)$$

وإن F_j المقابلة لأكبر قيمة كامنة λ_j تدعى بالمركبة الرئيسة الأولى، وتقوم بتفسير أكبر نسبة من التباين الكلي في مجموعة البيانات المعيارية Z_j. وأن مجموع التباين الكلي: $\sum_{j=1}^{K}\lambda_j = K$.

ولغرض تقليص عدد التراكيب (الأبعاد) التي ستستعمل في تفسير الظاهرة المدروسة، لايمكن إستعمال جميع المركبات الرئيسة، ولكن يمكن الأعتماد على بعض قواعد الأختيار. ومن اللافت للنظر أنه لايوجد إتفاق أو قبول عام بشأن أسلوب الأختيار هذا. فمن الباحثين من يفضل أختيار القيم الذاتية التي تزيد على الواحد الصحيح لتؤخذ بنظر الأعتبار[6]. بينما يقترح آخرون أن المركبات المختارة يجب أن تساهم في تفسير 70% أو أكثر من التباين الكلي، بمعنى آخر؛ يتم إختيار المركبات الرئيسة الأربع أو الخمس الأولى ذات الأسهام الأكبر وبشرط: $\frac{\sum_{j=1}^{r}\lambda_j}{K} \geq 0.70$ ، $r < k$ ، للمركبات الأربع أو الخمس الأولى[5].

وهناك مقترح آخر، بأن تكون النسبة المتراكمة من التباين الكلي ماتعادل 75% الى 80% للمركبات الخمسة أو الستة الأولى من المركبات الرئيسة. والجدول الآتي يبين ملخص تحليل المركبات الرئيسة بإستعمال مصفوفة الأرتباط[5].

جدول (1) يبين تحليل المركبات الرئيسة الأعتيادية بالأستناد الى مصفوفة الأرتباطات

المتغيرات	المركبات				مجموع مربعات عناصر الصف
	F_1	F_2	…	F_k	
Z_1	$a_{11}\sqrt{\lambda_1}$	$a_{12}\sqrt{\lambda_2}$	…	$a_{1K}\sqrt{\lambda_K}$	1
Z_2	$a_{21}\sqrt{\lambda_1}$	$a_{22}\sqrt{\lambda_2}$	…	$a_{2K}\sqrt{\lambda_K}$	1
. . .	. . .	. . .	. . .	. . .	. . .
Z_k	$a_{k1}\sqrt{\lambda_1}$	$a_{K2}\sqrt{\lambda_2}$	…	$a_{KK}\sqrt{\lambda_K}$	1
مجموع مربعات عناصر العمود (القيم الذاتية)	λ_1	λ_2	…	λ_K	
نسبة مساهمة كل مركبة في تفسير التباين الكلي	$\frac{\lambda_1}{K}$	$\frac{\lambda_2}{K}$	…	$\frac{\lambda_K}{K}$	
النسبة التجميعية	$\frac{\lambda_1}{K}$	$\frac{\lambda_1+\lambda_2}{K}$	…	1	

2.2 طريقة تقليص وإختيار المركبات الرئيسة المطلقة الصغرى

يمكن القول أن المركبات الرئيسة F_j تقوم بتعظيم التباين المتمثل بالمعادلة أدناه:

$$a'_j * R * a'_j \qquad \ldots(5)$$

وبشرط:

$$a'_j\, a_j = 1 \quad ؛ \quad \text{وعند } (j \geq 2)\quad a'_h\, a_j = 0 \quad \text{لكل } (h<j) \qquad \ldots(6)$$

سيكون هناك تعظيم لدالة التباين المعرف على المركبات الرئيسة، بإضافة القيد: $\sum_{i=1}^{K} |a_{ji}| \leq t$ لمعلمة t معينة على المعادلة .(6)

إذ أن: t يدعى بمعلمة التناوب (أو التناغم) Tuning Parameter ، حيث: $1 < t < \sqrt{p}$

وسيكون:

$$F_j^* = \max\left(a'_j\, Ra_j + \theta \sum_{i=1}^{K} |a_{ji}| \right) \qquad \ldots(7)$$

أذ أن: t تمثل معلمة التناوب tuning parameter توازن بين مقدار التباين والتحيز.

وأن: θ معامل ضرب مناسب يجب أختياره بعناية.

$\sum_{i=1}^{K} |a_{ji}|$: ستدعى معيار من النوع L_1 لأنه يعتمد على تصغير القيم المطلقة.

إذ سيقوم هذا القيد من خلال طبيعته معيار$-L_1$ بتقليص بعض المعاملات ليساويها أخيراً بالصفر، وفي مقابل ذلك سنحصل كسباً في دقة المقدرات من خلال التحكّم المتوازن بين مقدار التحيز والتباين.

ويمكن إيجاد مركبات رئيسة F_j تقوم بدورها بإيجاد مجموعة جزئية من المتغيرات الرئيسة وسيرمز لها F_j^* ، وأثبتت هذه الطريقة جدارة وكفاءة[1] .

وتجدر الأشارة الى أنه يجب المحاولة مع أكثر من قيمة لـ (t) للوصول للحل الأمثل. إذ لايوجد أسلوب قاطع ومحدّد لأختيار قيمة (t) المثلى، بل يتطلب الأمر إستعمال مبدأ المحاولة والخطأ ولكن ضمن المدى $1 < t < \sqrt{K}$)[6]).

3.2 طريقة تقليص بقيد جزاء إضافي

سيكون هناك تعظيم لدالة التباين المعرف على المركبات الرئيسة، بإضافة القيدين: $\sum_{i=1}^{K} |a_{ji}| \leq t$ و $\sum_{i=1}^{K} a_{ji}^2 \leq t$ للمعلمة t، ويتم الحصول على:

$$F_j^{**} = \max\left(a'_j\, Ra_j + \theta \left\{ \sum_{i=1}^{K} |a_{ji}| + \sum_{i=1}^{K} a_{ji}^2 \right\} \right) \qquad \ldots(8)$$

$\sum_{i=1}^{K} a_{ji}^2$: سيدعى بمعيار L_2 لأنه يعتمد تصغير قيم التربيعات.

وكما يلاحظ من أعلاه أن هناك عملية دمج وتهجين مابين تحقيق الأمثلية بمعيار L_1 والمتمثل بمجموع القيم المطلقة للمركبات، وبين معيار L_2 والمتمثل بمجموع القيم التربيعية لها. وسيتم تسميتها MLASS. وبالشرط نفسه على مدى المعلمة t في الطريقة LASS أعلاه.

3. الجانب التطبيقي

1.3 البيانات

إرتكز البحث على بيانات أستحصلت من إحدى الدراسات في المجال الأداري، شملت عينة من الموظفين بحجم (n=50). إذ كان المهم قياس مقدار وقوة القيادة الأستراتيجية Y (متغير الأستجابة التابع) والتي تتأثر تبعاً لبعض المتغيرات التوضيحية المهمة (K=9)، وكالآتي:

X_1: العمليات الجوهرية. X_2: القدرة على التعلّم. X_3: الموارد البشرية.

X_4: تكنولوجيا المعلومات. X_5: البحث والتطوير. X_6: الموارد الثمينة.

X_7: القدرات الأدارية والتنظيمية. X_8: ثقافة التداؤب. X_9: العلاقات وفهم حاجات الزبون.

وتم إجراء عملية التعيير Standardization لهذه المتغيرات، مع بقاوءها بالرموز X's نفسها.

2.3 النتائج

بإستعمال لغة MATLAB version 7.6 (Release 14) البرمجية في تحليل بيانات من خلال برنامج حاسوب خاص بالبحث تم كتابته ووضعه من قبل الباحثين لهذا الغرض ولكلا حالتي التحليل بالمركبات الرئيسة الأعتيادي والمقلص سواء بطريقة الـ (LASS) أو بالصيغة المعدلّة عنها (MLASS)، وتم الحصول على النتائج المبينة في الفقرات اللاحقة.

1.2.3 التحليل بالمركبات الرئيسة

بتطبيق ماجاء في الجدول (1) في الجانب النظري أعلاه، وحل المعادلتين (2) و(3) تم الحصول على قيم المركبات الرئيسة بعد تطبيق الصيغة في المعادلة (4)، مع أخذ أول خمسة مركبات فقط بنظر الأعتبار، وكالآتي:

جدول (2) يبين المركبات الرئيسة الخمسة الأولى الأعتيادية

Variable	PC_1	PC_2	PC_3	PC_4	PC_5
X_1	−0.1135	0.6855	−0.2850	0.2111	−0.1265
X_2	−0.3049	0.03008	0.6479	−0.2032	0.1191
X_3	−0.4138	−0.1979	0.1100	−0.3602	0.1902
X_4	−0.1506	0.5875	0.3851	−0.1507	−0.2948
X_5	−0.2853	0.1114	0.1806	0.6477	0.5800
X_6	−0.4555	0.0103	−0.3203	−0.1703	0.0479
X_7	−0.3881	0.1135	−0.4453	−0.2996	0.2143
X_8	−0.3588	−0.2810	−0.0529	0.4430	−0.2860
X_9	−0.3634	−0.2014	0.0482	0.1648	−0.6167
Variance (%)	39.0229	15.2457	11.0373	9.9192	7.9540
Cumulative Variance(%)	39.0229	54.2686	65.3059	75.2251	83.1791

يلخص الجدول (2) أعلاه نتائج تحليل المركبات الرئيسة بعد إختيار أول (أو أكبر) خمسة مركبات من ضمن التسعة الكلية، وقد بلغت النسبة المئوية لما تفسره من مجموع التباين الكلي تساوي (% 83.1791). والجدول (3) أدناه يوضح نسبة المساهمة المئوية من تفسير التبياين الكلي لطريقتي LASS و MLASS ولقيم t=1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00) أي بزيادة منتظمة قدرها (0.25).

جدول (3)

t	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
PC_{LASS}	69.2110	67.0549	73.0743	69.4445	65.3963	67.2633	67.6384	71.5241	70.4125
PC_{MLASS}	65.9677	68.3673	71.3481	68.6566	70.2764	70.7169	70.5164	73.6226	73.0736

إن الجداول رقم (4)، (5) و (6) جاءت بعد أن تم تطبيق طريقة التقليص على المركبات الرئيسة بالقيد المطلق فقط LASS وبالصيغة المذكورة في المعادلة (7)، والجداول رقم (7)، (8) و (9) جاءت بعد أن تم تطبيق طريقة التقليص على المركبات الرئيسة بالقيد المطلق والقيد التربيعي MLASS وبالصيغة المذكورة في المعادلة (8)، ، وكما في أدناه.

جدول (4) يبين المركبات الرئيسة الخمسة الأولى بطريقة LASS

t = 1.50						
Variable	PC_1	PC_2	PC_3	PC_4	PC_5	$\sum_{j=1}^{5} \lvert a_j \rvert$
X_1	0.000	0.121	0.904	0.000	0.000	1.025
X_2	−0.071	−0.291	0.000	1.000	0.000	1.362
X_3	−0.435	0.146	0.000	0.000	0.000	0.581
X_4	0.000	−0.002	0.428	0.000	0.000	0.430
X_5	−0.115	0.024	0.000	0.000	1.000	1.139
X_6	−0.823	−0.051	0.000	0.000	0.000	0.874
X_7	−0.124	−0.500	0.000	0.000	0.000	0.624
X_8	−0.149	−0.736	0.000	0.000	0.000	0.885
X_9	−0.279	−0.291	0.000	0.000	0.000	0.570
Variance (%)	24.8316	15.4024	12.6627	10.0888	10.0888	−
Cumulative Variance(%)	24.8316	40.2340	52.8967	62.9855	73.0743	−

جدول (5) بين المركبات الرئيسة الخمسة الأولى بطريقة LASS

t = 2.75						
Variable	PC_1	PC_2	PC_3	PC_4	PC_5	$\sum_{j=1}^{5}\lvert a_j\rvert$
X_1	0.000	−0.04	0.004	0.000	0.000	0.044
X_2	−0.071	−0.008	0.000	0.000	1.000	1.079
X_3	−0.435	0.002	1.000	0.000	0.000	1.437
X_4	0.000	−0.007	0.000	1.000	0.000	1.007
X_5	−0.115	−0.035	0.000	0.000	0.000	0.150
X_6	−0.823	−0.050	0.009	0.000	0.000	0.882
X_7	−0.124	0.005	0.000	0.000	0.000	0.129
X_8	−0.149	0.024	0.016	0.000	0.000	0.189
X_9	−0.279	−0.997	0.016	0.000	0.000	1.292
Variance (%)	26.8956	11.4307	11.3434	10.9273	10.9271	−
Cumulative Variance(%)	26.8956	38.3263	49.6697	60.5970	71.5241	−

جدول (6) يبين المركبات الرئيسة الخمسة الأولى بطريقة LASS

t = 3.00						
Variable	PC_1	PC_2	PC_3	PC_4	PC_5	$\sum_{j=1}^{5}\lvert a_j\rvert$
X_1	0.000	−0.003	0.124	0.000	0.000	0.127
X_2	−0.101	0.007	−0.076	1.000	0.000	1.184
X_3	−0.425	−0.035	0.010	0.000	0.000	0.470
X_4	0.000	0.000	−0.001	0.000	0.000	0.001
X_5	0.000	0.006	0.987	0.000	1.000	1.993
X_6	−0.822	0.951	−0.057	0.000	0.000	1.830
X_7	−0.151	0.307	0.003	0.000	0.000	0.461
X_8	−0.191	−0.013	0.019	0.000	0.000	0.223
X_9	−0.272	−0.009	−0.041	0.000	0.000	0.322
Variance (%)	25.6105	14.0673	10.3734	10.3734	9.9878	−
Cumulative Variance(%)	25.6105	39.6778	50.0513	60.4247	70.4125	−

جدول (7) يبين المركبات الرئيسة الخمسة الأولى بطريقة MLASS

t = 1.50

Variable	PC_1	PC_2	PC_3	PC_4	PC_5	$\sum_{j=1}^{5}\lvert a_j\rvert+\sum_{j=1}^{5}a_j^{2}$
X_1	0.0000	0.0491	0.8581	0.2650	0.0000	0.9905
X_2	−0.2050	−0.4835	0.0000	0.1349	1.0000	1.5586
X_3	−0.4597	−0.3139	0.0000	0.0041	0.0000	0.5437
X_4	0.0000	−0.7450	0.5134	−0.0069	0.0000	1.0419
X_5	−0.1686	−0.1517	0.0000	0.3909	0.0000	0.4577
X_6	−0.6032	0.0674	0.0000	−0.7039	0.0000	1.1191
X_7	−0.3803	−0.1861	0.0000	−0.3101	0.0000	0.5759
X_8	−0.3106	0.1081	0.0000	−0.0067	0.0000	0.2668
X_9	−0.3363	0.1902	0.0000	−0.4084	0.0000	0.6254
Variance (%)	29.0233	11.6638	11.3664	10.4881	8.8065	---
Cumulative Variance(%)	29.0233	40.6871	52.0535	62.5416	71.3481	---

جدول (8) يبين المركبات الرئيسة الخمسة الأولى بطريقة MLASS

t = 2.75

Variable	PC_1	PC_2	PC_3	PC_4	PC_5	$\sum_{j=1}^{5}\lvert a_j\rvert+\sum_{j=1}^{5}a_j^{2}$
X_1	0.0000	0.0887	0.8580	0.0316	0.0000	0.8616
X_2	−0.2050	0.3590	0.0000	−0.0627	0.0000	0.4007
X_3	−0.4597	−0.3113	0.0000	0.0105	0.0000	0.5449
X_4	0.0000	0.17122	0.5135	−0.0052	0.0000	0.4914
X_5	−0.1685	−0.0505	0.0000	−0.0259	1.0000	1.1382
X_6	−0.6033	−0.2347	0.0000	−0.0638	0.0000	0.6624
X_7	−0.3802	−0.3430	0.0000	0.0271	0.0000	0.5066
X_8	−0.3106	−0.6792	0.0000	−0.1190	0.0000	0.8403
X_9	−0.3363	−0.3164	0.0000	−0.9875	0.0000	1.4142
Variance (%)	27.9057	15.9472	10.9297	10.3719	8.4681	---
Cumulative Variance(%)	27.9057	43.8529	54.7826	65.1545	73.6226	---

جدول (9) يبين المركبات الرئيسة الخمسة الأولى بطريقة MLASS

Variable	PC_1	PC_2	PC_3	PC_4	PC_5	t = 3.00 $\sum_{j=1}^{5}\lvert a_j\rvert+\sum_{j=1}^{5}a_j^{\,2}$
X_1	0.0000	0.0405	0.3487	−0.1661	0.0000	0.3530
X_2	0−0.205	−0.0764	0.4306	−0.4047	0.0000	0.7568
X_3	−0.4597	0.3544	−0.1347	−0.4628	0.0000	0.9904
X_4	0.0000	−0.1012	0.0757	0.0346	0.0000	0.1143
X_5	−0.1685	0.0435	0.1914	−0.0439	1.0000	1.2580
X_6	−0.6032	0.5325	0.7444	−0.6516	0.0000	2.0789
X_7	−0.3803	0.7475	0.2566	0.1573	0.0000	1.1678
X_8	−0.3106	0.1111	−0.062	0.0008	0.0000	0.2985
X_9	−0.3364	0.0026	0.0923	0.3764	0.0000	0.5355
Variance (%)	25.7382	17.045	13.9425	8.5378	7.8101	-----
Cumulative Variance(%)	25.7382	42.7832	56.7257	65.2635	73.0736	-----

4. الأستنتاجات والتوصيات

يمكن ملاحظة بعض الأستنتاجات الواضحة التي أفرزها الجانب التطبيقي السابق من خلال النتائج المجدولة التي تم إستعراضها، وتلخيصها بالآتي:

1. نجحت الطريقة المقترحة MLASS بإعطاء نتائج جيدة عند (t= 1.5 , 2.75, 3.00)، إذ بلغت نسب المساهمة في تفسير التباين الكلي للمركبات (71.3481 %, 73.6226 %, 73.0736) على التوالي مع الأختزال الواضح في قيم المركبات في الجداول (7, 8, 9) على التوالي. وكانت نتائج طريقة LASS لتقليص المقدرات قد سجلت أفضلية عند (t= 1.5 , 2.75, 3.00) ، إذ بلغت القيم لنسبة التباين الكلي المفسّر (70.4125 %, 71.5241 %, 73.0743 %) على التوالي، ناهيك عن الأختزال الواضح في قيم المركبات في الجداول (4, 5, 6). هذا وكانت قيمة نسبة التباين الكلي المفسّر هي (83.18 %) للمركبات الرئيسة الأعتيادية.
2. تم الحصول على أعلى إختزال في طريقتي التقليص LASS و MLASS للمركبة الخامسة بالأستناد الى نسبة مساهمة المركبات الرئيسة الى المجموع الكلي.
3. لم يتحقق شرط التقليص دائماً، وبخاصة عند الحالات التي لم تحقق فيها (t) الأمثلية، في طريقة MLASS وعند (t=1.50) للجدول رقم (7) في الصف الثاني لعمود القيد.
4. يوصي الباحثون بإمكانية تطبيق طريقة التقليص المذكورة في هذا البحث ولكن بإتباع أسلوب آلي (Automatic) ضمن معيار تكراري (Iterative) معين يخصص لهذا الغرض لأختيار قيمة (t) التي تحقق الأمثلية في شرط التقليص، بدلاً من الأسلوب الشخصي (Subjective) من خلال الزيادة الثابتة على قيم (t).

المصادر

[1] علي، د.عمرعبدالمحسن؛ المهنا، د.فراس أحمد محمد؛ (2009)؛ "حول تقليص تقدير المركبات الرئيسة مع التطبيق"؛ المؤتمر العلمي الثاني للرياضيات – الأحصاء والمعلوماتية، كلية علوم الحاسبات والرياضيات، جامعة الموصل.

[2] Jeffers, J.; (1967); "Two Case Studies in the application of Principal Components"; Applied Statistics, Vol. 16, pp. 225-236.

[3] Morrison, D.F.; (1976); "Multivariate Statistical Methods"; McGraw-Hill, New-York.

[4] Tibshirani, R.; (1996); "Regression Shrinkage and Selection via the LASS"; Journal of the Royal Statistical Society Series B Methodological, Vol. 58 No.(1), pp. 267 – 288.

[5] Tim, Neil H.; (2002); "Applied Multivariate Analysis"; Springer - Verlag, New - York.

[6] Zou, H.; Hastie, T. and Tibshirani, R.; (2004); "Sparse Principal Component Analysis"; Tech. Rep. Statistics Department, Stanford University.

"Modified LASS Method Suggestion as an additional Penalty on Principal Components Estimation – with Application-"

Abstract

This research deals with a shrinking method concernes with the principal components similar to that one which used in the multiple regression "Least Absolute Shrinkage and Selection: LASS". The goal here is to make an uncorrelated linear combinations from only a subset of explanatory variables that may have a multicollinearity problem instead taking the whole number say, (K) of them. This shrinkage will force some coefficients to equal zero, after making some restriction on them by some "tuning parameter" say, (t) which balances the bias and variance amount from side, and doesn't exceed the acceptable percent explained variance of these components. This had been shown by MSE criterion in the regression case and the percent explained variance in the principal components case.