به نام خدا

تکنیک آنالیز اجزای اصلی
(PCA)

گردآوری

شهرام کبیری

کلمات کلیدی

PCA، بردار های ویژه، مقادیر ویژه

چکیده

این مقاله تشریح مفهوم ***PCA*** جمع آوری شده است. ***PCA*** یک شگرد آماری مفید است که در زمینه هایی مثل تشخیص چهره و پردازش تصویر کاربرد دارد ، و یک تکنیک معمول در یافتن الگو در داده های با ابعاد زیاد می باشد .

۱. مقدمه

پیشرفتهای به وجود آمده در جمع آوری داده ها و قابلیتهای ذخیره سازی در طی دهه های اخیر باعث شده در بسیاری از علوم با حجم بزرگی از اطلاعات روبرو شویم .محققان در زمینه های مختلف مانند مهندسی ،ستاره شناسی،زیست شناسی و اقتصاد هر روز با مشاهدات بیشتر و بیشتری روبرو می شوند در مقایسه با بسترهای داده ای قدیمی و کوچکتر بسترهای داده ای امروزی چالشهای جدیدی در تحلیل داده ها به وجود آورده اند .روشهای آماری سنتی به دو دلیل امروزه کارایی خود را از دست داده اند

علت اول: افزایش تعداد مشاهدات۱ است

علت دوم : که از اهمیت بالایی برخوردار است افزایش تعداد متغیرهای مربوط به یک مشاهده است

تعداد متغیرهایی که باید برای هر مشاهده اندازه گیری شود ابعاد داده نامیده می شود عبارت متغیر۲ بیشتر در آمار استفاده می شود در حالی که در علوم کامپیوتر و یادگیری ماشین بیشتر از عبارت ویژگی۳ و یا صفت۴ استفاده می شود.

بسترهای داده ای که دارای ابعاد زیادی هستند علیرغم فرصت هایی که به وجود می آورند چالشهای محاسباتی زیادی را ایجاد می کنند یکی از مشکلات داده های با ابعاد زیاد این است که در بیشتر مواقع تمام ویژگیهای داده ها برای یافتن دانشی که در داده ها نهفته است مهم و حیاتی نیستند.به همین دلیل در بسیاری از زمینه ها کاهش ابعاد داده ها یکی از مباحث قابل توجه باقی مانده است.

۲. روش های کاهش ابعاد

روشهای کاهش ابعاد داده به دو دسته تقسیم می شوند:

روشهای مبتنی بر استخراج ویژگی:این روشها یک فضای چند بعدی را به یک فضایی با ابعاد کمتر نگاشت می کنند در واقع با ترکیب مقادیر ویژگیهای موجود تعداد کمتری ویژگی به وجود می آورند به طوری که این ویژگیها دارای تمام (یا بخش اعظمی از) اطلاعات موجود در ویژگیهای اولیه باشند این

۱- Observation

۲- Variable

۳- Feature

۴- Attribute

روشها به دو دسته خطی و غیر خطی تقسیم می شوند که موضوع مورد بحث در همین حوزه قرار می گیرد.

روشهای مبتنی بر انتخاب ویژگی : این روشها سعی می کنندبا انتخاب زیر مجموعه ای از ویژگیهای اولیه ابعاد داده ها را کاهش دهند.

در این فصل کمتر به اثباتهای ریاضی پرداخته شده و بیشتر به مفاهیم و کاربرد روش ها توجه شده است در تهیه مطالب این فصل سعی شده است با ارائه مثالهای مناسب خواننده در درک بهتر مفاهیم یاری شود

۲-۱. روشهای مبتنی بر استخراج ویژگی:

همانطور که اشاره شد روشهای مبتنی بر استخراج ویژگی ،یک فضای چند بعدی را به یک فضایی با ابعادکمتر نگاشت می دهند.این روشها بر دو دسته خطی و غیر خطی تقسیم می شوند روشهای خطی که ساده ترند و فهم آنها را حتتر است به دنبال یافتن یک زیر فضای تخت عمومی[۱] هستند اما روشهای غیر خطی که مشکلترند و تحلیل آنها سخت تر است به دنبال یافتن یک زیر فضای تخت محلی[۲] هستند.

از روشهای خطی می توان به ***FA***و***PCA*** ، ***DWT*** ، ***DFT*** اشاره کرد.

از روشهای غیر خطی می توان به موارد زیر اشاره کرد:

- ***Principal Curves***
- ***Self Organizing Maps***
- ***Vector Quantization***
- ***Genetic and Evolutionary Algorithms***
- ***Regression***

مسئله کاهش ابعاد داده را بطور ریاضی می توان به اینصورت بیان کرد :

یک متغیر تصادفی $p-$ بعدی $x=(x_1,x_2,...,x_p)^T$ داریم می خواهیم متغییر $k-$ بعدی $s=(s_1,s_2,...,s_k)^T$را به گونه ای پیدا کنیم که اولاً $K\le p$ و ثانیاً s محتویاتی که در x وجود دارد را براساس معیاری خاص دارا باشد روشهای خطی سعی می کنند هر یک از این k مؤلفه را از ترکیب خطی p مؤلفهٔ اولیه بدست آورند(***T*** به معنای ترانهاده است).

۱- Global Flat Subspace

۲- Locally Flat Subspace

$$s = WX \Rightarrow \begin{matrix} s_i = w_{i,1}x_1 + w_{i,2}x_2 + \ldots + w_{i,p}x_p \\ i = 1,2,\ldots k \end{matrix}$$

که $W_{k\times p}$ ماتریسهای وزنهای نگاشت خطی می باشد.

تکنیک ***PCA*** بهترین روش برای کاهش ابعاد داده به صورت خطی می باشد یعنی با حـذف ضـرایب کـم اهمیت به دست آمده از این تبدیل اطلاعات از دست رفته نسبت به روشهای دیگر کمتر است.

۳. تکنیک آنالیز اجزای اصلی[۳] (PCA)

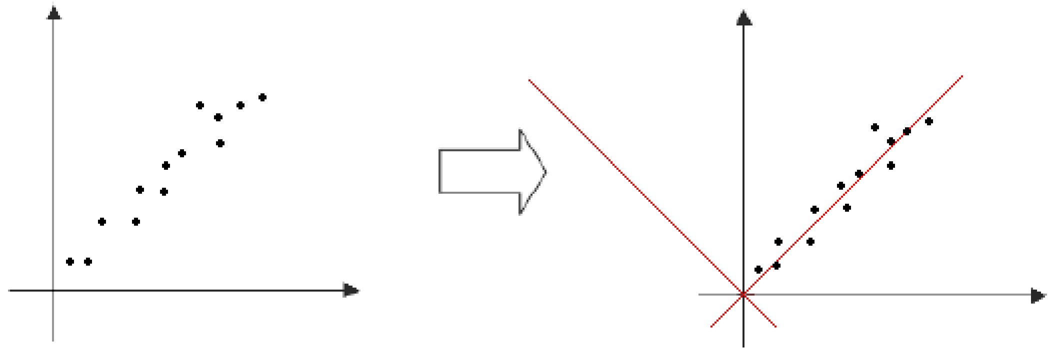
در این روش محورهای مختصات جدیدی برای داده ها تعریف شـده و داده هـا براسـاس ایـن محورهـای مختصات جدید بیان می شوند .اولین محورباید در جهتی قرار گیرد که واریانس داده ها ماکسـیمم شـود (یعنی در جهتی که پراکندگی داده هابیشتر است) دومین محور باید عمود بر محور اول به گونه ای قـرار گیرد که واریانس داده ها ماکسیمم شود به همین ترتیب محورهای بعدی عمود بـر تمـامی محـور هـای قبلی به گونه ای قرار می گیرند که داده ها در آن جهت دارای بیشترین پراکندگی باشند در شـکل ۱-۵ این مطلب برای داده های دوبعدی که ملموس تر است نشان داده شده است.

روش ***PCA*** به نامهای دیگری نیز معروف است مانند:

Karhunen Loeve Transform(KLT)
Hotelling Transform
Empirical Orthogonal Function(EOF)

قبل از اینکه به جزئیات این روش بپردازیم ابتدا مفاهیم ریاضـی و آمـاری مـرتبط بـا ایـن روش را بطـور مختصر بیان می کنیم این مفاهیم شامل انحراف از معیار اسـتاندارد،کوواریانس ،بردارهـای ویـژه و مقـادیر ویژه می باشند.

۳- Principal Component Analysis



شکل۱ :محورهای جدید باتوجه به بردارهای ویژه در جهت پر تراکم ترین نقاط قرار دارند

۴. مفاهیم مقدماتی مورد نیاز در PCA:

۴-۱. مفاهیم آماری:

فرض کنید X رشته ای از مقادیر است میانگین این مقادیر از رابطهٔ زیر به دست می آید:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

واریانس نیز از رابطهٔ زیر محاسبه می شود.

$$V(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

علت اینکه در رابطه فوق از عبارت $n-1$ بجای n استفاده شده اینست که فرض شده تمام مقادیر موجود نیست بلکه تعدادی از این مقادیر انتخاب شده اند و در X قرار گرفته اند یعنی X مجموعه نمونه است ونه کل داده ها ، با این فرض اگر از $n-1$ در رابطه فوق استفاده شود به واریانس داده های واقعی نزدیکتر خواهد بود نسبت به اینکه از n استفاده شود .

عیب واریانس آنست که آنراحتی باخود صفت هم نمی توان مقایسه کرد زیرا اگر فرض کنیم صفت برحسب متر باشد، $X_i - \overline{X}$ نیز برحسب متر می باشد ولی $(X_i - \overline{X})^2$ بر حسب متر مربع است لذا قابل مقایسه نمی باشد و این عیب واریانس است برای رفع آن از جذر مثبت واریانس ، انحراف معیار یا انحراف استاندارد استفاده می کنیم.

$$\sigma = \sqrt{V(X)}$$

معیارهایی که در بالا ذکر شد فقط اطلاعات مربوط به یک بعد را ارائه می کنند و دانشی در مورد ارتباط بین ابعاد مختلف را به ما نمی دهند با استفاده از کواریانس(همپراش) می توانیم ارتباط بین ابعاد مختلف داد ها را پیدا کنیم .

فرض کنید یک رشته دیگر از اعداد داریم که با Yنشان می دهیم کواریانس بین X,Yاز رابطه زیر بدست می آید:

$$Cov(X,Y)=\frac{1}{n-1}\sum_i (X_i-\overline{X})(Y_i-\overline{Y})$$

$Cov(X,Y)$برحسب اینکه مثبت یا منفی باشد نشان می دهد که X,Y بطور متوسط در یک جهت یا در دو جهت مخالف تغییر می کنند اگر مقادیر بزرگ Xیعنی مقادیری که از میانگین بزرگتر هستند ،با مقادیربزرگ Y و یا مقادیر کوچک X با مقادیر کوچک Y مربوط شوند کواریانس مثبت است واگر مقادیربزرگ X با مقادیر کوچک Yو یا مقادیر کوچک X با مقادیر بزرگ Y مربوط شوند کواریانس منفی خواهد بود.

یا به زبان ساده تر:

اگر $Cov(X,Y)>0$ آنگاه X,Y مشابه هم تغییر می کنند.

اگر $Cov(X,Y)<0$ آنگاه X,Y متضاد هم تغییر می کنند.

اگر $Cov(X,Y)=0$ نمی توان نتیجه گرفت که X,Y مستقل هستند (بجز حالتی که X,Yهر دو نرمال باشند)

کواریانس بین تمامی ابعاد داده ها را می توان دو به دو محاسبه کرد و در یک ماتریس ذخیره کرد به این ماتریس ماتریس کواریانس می گویند .ماتریس کواریانس یک ماتریس مربعی متقارن است مثلاً اگر سه بعد به نامهای X,Y,Z داشته باشیم ماتریس کواریانس آنها برابر است با:

$$C=\begin{pmatrix} \operatorname{cov}(x,x) & \operatorname{cov}(x,y) & \operatorname{cov}(x,z) \\ \operatorname{cov}(y,x) & \operatorname{cov}(y,y) & \operatorname{cov}(y,z) \\ \operatorname{cov}(z,x) & \operatorname{cov}(z,y) & \operatorname{cov}(z,z) \end{pmatrix}$$

علت متقارن بودن ماتریس کواریانس اینست که:

$$Cov(X,Y)=\frac{1}{n-1}\sum_i (X_i-\overline{X})(Y_i-\overline{Y})=Cov(Y,X)$$

۴-۱-۱. مفاهیم جبر ماتریسها

در این بخش مفهوم بردار ویژه و مقادیر ویژه را بیان می کنیم همانطور که می دانید برای اینکه بتوان دو ماتریس را در هم ضرب کرد باید تعداد ستونهای ماتریس اول برابر تعداد سطرهای ماتریس دوم باشد.بردارهای ویژه نوع خاصی از ضرب ماتریسها را ارائه می کنند به مثالهای زیر توجه کنید:

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

در مثال اول بردار بدست آمده مضرب صحیحی از بردار اولیه نیست اما در مثال دوم بردار به دست آمده چهار برابر بردار اولیه می باشد ماتریس ۲×۲ی که در این دو بردار ضرب کرده ایم را می توان یک ماتریس تبدیل۱ در نظر گرفت که با ضرب آن در یک بردار می توان اندازه و راستای آن بردار را تغییر داد در میان تمام بردارهایی که می توان ماتریس بردار را در آنهاضرب کرد بردارهایی وجود دارند که پس از تبدیل راستای آنها تغییر نمی کند و فقط اندازه آنها تغییر می کند مانند بردار $[3 \quad 2]^T$ در مثال مذکور ، این بردارها را بردارهای ویژه می نامند برای هر بردار ویژه یک مقدار ویژه نیز وجود دارد که بیان می کند اندازه آن بردار (وتمام بردارهای دیگر که در راستای آن بردار هستند) پس از تبدیل چند برابر خواهد شد در مثال فوق مقدار ویژه برای بردار $[3 \quad 2]^T$ و البته تمام بردارهای هم راستا با آن مانند $[6 \quad 4]^T$ برابر ۴ خواهد بود.(علامت ***T*** به معنای ترانهاده است)

$$2 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 24 \\ 16 \end{pmatrix} = 4 \times \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

بردارهای ویژه و مقدارهای ویژه فقط برای ماتریسهای مربعی معنی پیدا می کنند یک ماتریس $n \times n$ می تواند دارای n بردار ویژه باشد به منظور استاندارد کردن بردارهای ویژه کلیه عناصرموجوددر بردار ویژه را بر طول بردار تقسیم می کنیم .مثلاً برای بردار $[3 \quad 2]^T$ داریم:

۱- Transformation Matrix

$$\sqrt{(3^2+2^2)} = \sqrt{13} \qquad \begin{pmatrix} 3 \\ 2 \end{pmatrix} \div \sqrt{13} = \begin{pmatrix} 3/\sqrt{13} \\ 2/\sqrt{13} \end{pmatrix}$$

ویژگی مهم بردارهای ویژه آنست که برهم عمودند .

۴-۱-۲. نحوه محاسبه مقدارهای ویژه و برادرهای ویژه یک ماتریس مربعی:

تبدیل خطی زیر را درنظر بگیرید:

$$y = AX = \lambda X \qquad (1)$$

(A حقیقی است)

حال نشان می دهیم که هر ماتریس مربعی n سطری حداقل ۱ و حداکثر n مقدار ویژه متمایز (حقیقی یا مختلط) دارد.

برای این منظور (۱)را باز می کنیم:

$$\begin{aligned} &a_{11}x_1 + \ldots + a_{1n}x_n = \lambda x_1 \\ &a_{21}x_1 + \ldots + a_{2n}x_n = \lambda x_2 \\ &. \\ &. \qquad\qquad\qquad\qquad (2) \\ &. \\ &a_{n1}x_1 + \ldots + a_{nn}x_n = \lambda x_n \end{aligned}$$

پس از انتقال جملات سمت راست به چپ خواهیم داشت:

$$\begin{aligned} &(a_{11}-\lambda)x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\ &a_{21}x_1 + (a_{22}-\lambda)x_2 + \ldots + a_{2n}x_n = 0 \\ &. \\ &. \qquad\qquad\qquad\qquad (3) \\ &. \\ &a_{n1}x_1 + a_{n2}x_2 + \ldots + (a_{nn}-\lambda)x_n = 0 \end{aligned}$$

وبا نماد ماتریسی داریم:

$$(A - \lambda I) = 0$$

بنا به قضیه کرامر دستگاه همگن معادلات خطی تنها و تنها وقتی جواب غیر بدیهی دارد که دترمینان ضرایب آن صفر باشد پس داریم:

$$D(\lambda)=\det(A-\lambda I)=\begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ . & . & & \\ . & . & & \\ . & . & & \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix}=0 \qquad (4)$$

$D(\lambda)$ را دترمینان مشخصه و رابطه بالا را معادله مشخصه متناظر با ماتریس A می نامند با بسط $D(\lambda)$ یک چند جملهٔ درجه n بر حسب λ به دست می آوریم این چند جمله ای را چند جمله ای مشخصه متناظر با A نامند.

بنا به قضییه زیر را داریم:

مقادیر ویژه ماتریس مربعی $\boldsymbol{A}$ ریشه های معادلهٔ مشخصه متناظر با (۴) هستند.

مقدار ویژه ای را که ریشه مرتبه $\boldsymbol{m}$ ام چند جمله ای مشخصه است مقدار ویژه مرتبه $\boldsymbol{m}$ ام ماتریس متناظر نامند.

به مجرد اینکه مقادیر ویژه تعیین گردیدند بردارهای ویژه متناظر را می توان از دستگاه (۳) به دست آورد چون دستگاه همگن است پیداست که هر گاه $\boldsymbol{x}$ بردار ویژه $\boldsymbol{A}$ باشد آنگاه $\boldsymbol{kx}$ نیز که درآن $\boldsymbol{k}$ مقداری ثابت و غیر صفر است بردار ویژه $\boldsymbol{A}$ متناظر با همان مقدار ویژه است.

مثال: محاسبه بردارهای ویژه استاندارد شده ماتریس $\boldsymbol{A}$:

$$A=\begin{pmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{pmatrix}$$

$$|A-\lambda I|=\begin{vmatrix} 0.6165-\lambda & 0.6154 \\ 0.6154 & 0.7165-\lambda \end{vmatrix}=\lambda^2-1.333\lambda+0.061=0$$

یافتن ریشهای معادله درجه دو با استفاده از نرم افزار ***MATLAB*** که همان مقادیر ویژه مورد نظر ما هستند:

```
>> p=[1 -1.333 0.061];
>> roots(p)
ans =
   1.28554947298257
   0.04745052701743
```

مقادیر ویژه ماتریس A

یافتن بردارهای ویژه با استفاده از مقادیر ویژه:

$$\Rightarrow\begin{bmatrix}0.6165 & 0.6154\\0.6154 & 0.7165\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=1.285\begin{bmatrix}x_1\\x_2\end{bmatrix}\Rightarrow\begin{cases}0.6165x_1+6154x_2=1.285x_1\Rightarrow 0.6154x_2=0.6685x_1\\0.6154x_1+0.7165x_2=1.285x_2\end{cases}$$

$$x_1=0.9205x_2\Rightarrow x_1=\begin{bmatrix}\dfrac{1}{\sqrt{1+0.9205^2}}\\\dfrac{0.9205}{\sqrt{1+0.9205}}\end{bmatrix}=\begin{bmatrix}0.7357\\0.6772\end{bmatrix}$$

$$\Rightarrow\begin{bmatrix}0.6165 & 0.6154\\0.6154 & 0.7165\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=0.047\begin{bmatrix}x_1\\x_2\end{bmatrix}\Rightarrow\begin{cases}0.6165x_1+6154x_2=0.047x_1\Rightarrow 0.6154x_2=-0.5695x_1\\0.6154x_1+0.7165x_2=0.047x_2\end{cases}$$

$$x_2=-0.9254x_1\Rightarrow x_2=\begin{bmatrix}\dfrac{-0.9254}{\sqrt{1+0.9254^2}}\\\dfrac{1}{\sqrt{1+0.9254}}\end{bmatrix}=\begin{bmatrix}-0.6769\\0.7360\end{bmatrix}$$

$$\Rightarrow\begin{bmatrix}0.7357 & -0.6772\\0.6772 & 0.7357\end{bmatrix}\qquad(5)$$

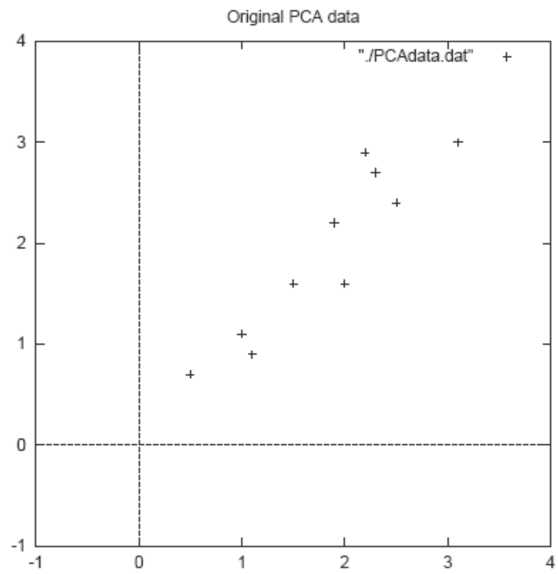
ماتریس (۵) بردارهای ویژه استاندارد شده ماتریس A می باشد.

۴-۲. الگوریتم PCA:

در این بخش الگوریتم ***PCA*** را با ذکر مثال توضیح می دهیم:

۴-۲-۱. مرحله اول انتخاب داده:

در این قسمت ما قصد داریم ***PCA*** را برروی یک مجموعه داده های دوبعدی اعمال کنیم این داده هـا در شکل ۲ نشان داده شده است.



(الف)

Data =

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

(ب)

شکل۲:(الف):نحوه پراکندگی داده ها(ب):مختصات داده ها

۴-۲-۲. مرحله دوم محاسبه ماتریس کواریانس:

ماتریس کواریانس را مطابق فرمول محاسبه ماتریس کواریانس بدست می آوریم.

داریم:

$$Cov(x,y)=\frac{1}{9}\sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y})$$

$$\bar{x}=\frac{1}{10}\sum_{i=1}^{10}x_i=1.81 \quad , \qquad \bar{y}=\frac{1}{10}\sum_{i=1}^{10}y_i=1.91$$

$$\Rightarrow Cov(x,y)=\frac{1}{9}\sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{9}((0.69\times0.49)+(-1.31\times(-1.21))+(0.39\times0.99)+$$
$$(0.09\times0.29)+(1.29\times1.09)+(0.49\times0.79)+(0.19\times(-0.31))+(-0.81\times(-0.81))+(-0.31\times(-0.31))+$$
$$(-0.71\times(-1.01)))=0.61544$$

مقادیرمحاسبه شده تنها برای عناصرروی قطرفرعی ماتریس کواریانس می باشد پس از محاسبه کامل

داریم:

$$cov=\begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

۴-۲-۳. مرحله سوم محاسبه بردارهای ویژه ومقادیر ویژه:

اکنون بردارهای ویژه و مقدارهای ویژه $Cov(x,y)$ را بدست می آوریم طبق قضایای جبر خطی یک ماتریس متقارن $n\times n$ دارای n بردار ویژه مستقل و n مقدار ویژه می باشد.

پیشتردر همین فصل ما بردارهای ویژه و مقادیر ویژه این ماتریس را محاسبه کردیم(۵) پس خواهیم داشت:

مقادیر ویژه:

$$eigenvalues = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

بردارهای ویژه:

$$eigenvectors = \begin{pmatrix} -.735178656 & -.677873399 \\ .677873399 & -.735178656 \end{pmatrix}$$

بردارهای ویژه بدست آمده در بخش قبل هم اندازه و هم راستای بردار ویژه فوق است.اما این دو بردار چه چیزی به ما می دهند ؟ راستای این بردار در شکل۳-۵.ب نشان داده شده است همانطور که می بینیدیکی از این دو بردار در جهتی قرار گرفته است که داده های در آن جهت بیشترین پراکندگی را دارند بردار دیگر نیز عمود بر بردار اول است و اما مقادیر ویژه چه چیزی را ارائه می دهند؟در این مثال برداری که در راستای بیشترین پراکندگی دادها قرار گرفته دارای مقدار ویژه ۱.۲۸۴ و بردار دیگر دارای مقدار ویژه ۰.۰۴۹ می باشد در واقع مقادیرویژه میزان پراکندگی داده ها در راستای بردار ویژه مربوطه را نشان می دهد می توان گفت بردار ویژه ای که دارای بزرگترین مقدار ویژه است مؤلفۀ اصلی داده های موجود است.

۴-۲-۴. مرحله چهارم انتخاب مؤلفه ها و ساختن Feature Vector

در این مرحله مفهوم کاهش ابعاد داده وارد می شود بردارهای ویژه ای که در مرحله قبل به دست آوردیم را براساس مقادیر ویژه آنها از بزرگ به کوچک مرتب می کنیم (توجه داشته باشید که مقادیر ویژه ماتریس کواریانس همگی بزرگتر یا مساوی صفر هستند)بدین ترتیب مؤلفه های داده ها از پر اهمیت به کم اهمیت مرتب می شوند در اینجا اگر بخواهیم ابعاد داده ها را کاهش دهیم می توانیم مؤلفه های کم اهمیت را حذف کنیم البته اینکار با از دست دادن مقدارکمی اطلاعات همراه است.

کاری که باید در این مرحله انجام دهیم ایجاد یک ***Feature Vector*** است که در واقع ماتریسی از بردارهامی باشد این ماتریس شامل آن بردارهای ویژگی ای است که ما می خواهیم آنها رانگه داریم.

$$FeatureVector = (eig_1\ eig_2\ eig_3\ ....\ eig_n)$$

اگر همهٔ بردارهای ویژگی را در این ماتریس قرار دهیم هیچ اطلاعاتی از دست نخواهد رفت و دوباره می توانیم دقیقاً همان داده های اولیه را بدست آوریم در ادامه مثال فوق ***Feature Vector*** را برابر با مقدار زیر در نظر می گیریم.

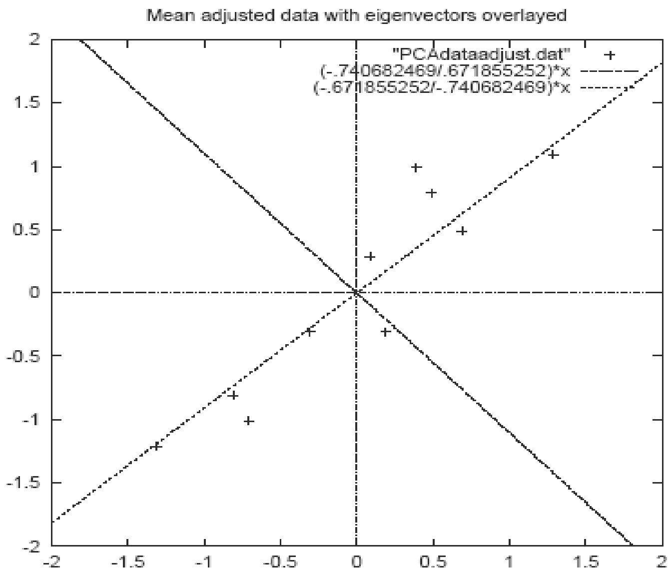
$$FeatureVector = \begin{pmatrix} -.677873399 \\ -.735178656 \end{pmatrix}$$

۴-۲-۵. مرحله پنجم بدست آوردن داده های جدید:

در آخرین مرحله از ***PCA*** فقط باید ترانهاده ***Feature Vector*** که در مرحله قبل به دست آوردیم را در ترانهاده داده های نرمال سازی شده ضرب کنیم.

$$FinalData = RowFeatureVector \times RowDataAdjust,$$

که ***RowFeatureVector*** ماتریسی است که بردارهای ویژه در سطرهای آن به ترتیب مقادیر ویژه از مقادیر زیاد به مقادیرکم قرار گرفته اند و ***RowDataAdjust*** ماتریسی شامل داده هایی است که میانگین هر بُعد ازآن کم شده در این ماتریس داده ها در ستونهای آن ذخیره شده و هر سطر آن مربوط به یک بُعد است در مثال مذکور بدلیل اینکه ما فقط یکی از بردارهای ویژه را انتخاب کردیم داده های بدست آمده ***PCA*** داده های یک بعدی می باشند (شکل۴).



(ب)

DataAdjust =

x	y
.69	.49
-1.31	-1.21
.39	.99
.09	.29
1.29	1.09
.49	.79
.19	-.31
-.81	-.81
-.31	-.31
-.71	-1.01

(الف)

شکل۳:(الف):مختصات داده های نرمالیزه شده (ب):محل قرار گیری محورهای جدید با توجه به بردارهای ویژه

x
-.827970186
1.77758033
-.992197494
-.274210416
-1.67580142
-.912949103
.0991094375
1.14457216
.438046137
1.22382056

شکل ۴: داده های بُعد ***x*** پس از اعمال تکنیک ***PCA***

با استفاده از رابطه زیر می توانیم مقادیری که از تبدیل ***PCA*** بدست آورده ایم را به داده های اولیه که مقدار میانگین از آنهاکم شده باز گردانیم.

$$FinalData = RowFeatureVector \times RowDataAdjust,$$

$$FinalData = RowFeatureVector \times RowDataAdjust) \times RowFeatureVector^{-1}$$

$$RowDataAdjust = RowFeatureVector^{-1} \times FinalData$$

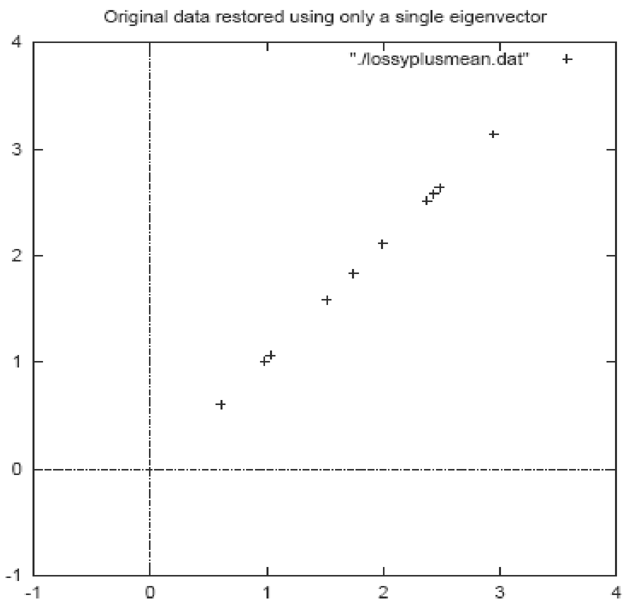
بدلیل اینکه ماتریس ***RowFeatureVector*** حاوی بردارهای ویژه یکه است معکوس آن با ترانهاده آن برابر است پس داریم:

$$RowDataAdjust = RowFeatureVector^T \times FinalData$$

با اضافه کردن میانگین، داده های اولیه را خواهیم داشت:

$$RowOriginalData = (RowFeatureVector^T \times FinalData) + OriginalMean$$

در شکل ۵-۵ داده هایی که پس از تبدیل ***PCA*** بازیابی شده اند را مشاهده می کنید همانطور که می بینید مقدار کمی از بین رفته است و باعث شده داده ها همگی در امتداد یک خط راست قرار گیرند.



شکل ۵ :مکان داد های بازیابی شده پس از اعمال تکنیک ***PCA***

در این بخش قصد داریم تکنیک ***PCA*** را بر روی ۶ ماتریس مشابه پروژه(شکل ۶-۵)اعمال کنیم البته در ابعاد کوچکتر تا با مراحل کار بهتر آشنا شویم:

مرحله اول انتخاب داده ها:

```
w1 =
   255   255   255   255   255
   255     0     0     0   255
   255     0   255     0   255
   255     0     0     0   255
   255   255   255   255   255

w2 =
   255   255   255   255   255
   255   255     0     0   255
   255   255     0     0   255
   255   255   255   255   255
   255   255   255   255   255

w3 =
   255   255   255   255   255
   255     0     0     0     0
     0   255   255   255     0
     0     0     0     0     0
   255   255   255   255   255

w4 =
   255   255   255   255   255
   255     0     0     0   255
   255     0   255     0   255
   255   255     0   255   255
   255   255   255   255   255

w5 =
   255   255   255   255   255
   255     0     0     0   255
   255     0     0     0   255
   255     0     0     0   255
   255   255   255   255   255

w6 =
   255   255   255   255   255
   255     0     0     0   255
   255     0   255     0   255
   255     0     0     0   255
   255   255   255   255   255
```

شکل ۶ : داده های ورودی جهت اعمال تکنیک ***PCA***

مرحلهٔ دوم ذخیره ۶ ماتریس مربوط به نمونه های عدد صفر در ماتریس *R*:

نکته قابل توجه اینست که نحوه ذخیره سازی ماتریسهای ***W1*** تا ***W6*** در ماتریس ***R*** به صورت ستونی می باشد در هر سطر ماتریس ***R*** مشخصات یکی از ماتریسهای w_i قرار دارد.

MATLAB Array Editor: R
Dec 27, 2007

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	255	255	255	255	255	255	0	0	0	255	255	0	255	0	255	255	0	0	0	255	255	255	255	255	255
2	255	255	255	255	255	255	255	255	255	255	255	0	0	255	255	255	0	0	255	255	255	255	255	255	255
3	255	255	0	0	255	255	0	255	0	255	255	0	255	0	255	255	0	255	0	255	255	0	0	0	255
4	255	255	255	255	255	255	0	0	255	255	255	0	255	0	255	255	0	0	255	255	255	255	255	255	255
5	255	255	255	255	255	255	0	0	0	255	255	0	0	0	255	255	0	0	0	255	255	255	255	255	255
6	255	255	255	255	255	255	0	0	0	255	255	0	255	0	255	255	0	0	0	255	255	255	255	255	255

جدول ۱ :ماتریس نگهداری مقادیر ورودی ماتریس های ***W1***و***W6***

$$R(i,:) = W_i$$

مرحلهٔ سوم محاسبه ماتریس کواریانس ***R*** و ذخیره آن در ماتریس ***CovR***:

MATLAB Array Editor: CovR
Nov 16, 2007

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335	0	0	0	-4335	2167.5	0	0	0	-10837.5	4335	0	0	10837.5	10837.5	10837.5	0
4	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335	0	0	0	-4335	2167.5	0	0	0	-10837.5	4335	0	0	10837.5	10837.5	10837.5	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	2167.5	2167.5	0	0	10837.5	8670	8670	0	0	0	-8670	10837.5	0	0	0	-2167.5	8670	0	0	2167.5	2167.5	2167.5	0
8	0	0	-8670	-8670	0	0	8670	17340	4335	0	0	0	-4335	8670	0	0	0	8670	4335	0	0	-8670	-8670	-8670	0
9	0	0	4335	4335	0	0	8670	4335	17340	0	0	0	-4335	8670	0	0	0	-4335	17340	0	0	4335	4335	4335	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	-4335	-4335	0	0	-8670	-4335	-4335	0	0	0	17340	-8670	0	0	0	4335	-4335	0	0	-4335	-4335	-4335	0
14	0	0	2167.5	2167.5	0	0	10837.5	8670	8670	0	0	0	-8670	10837.5	0	0	0	-2167.5	8670	0	0	2167.5	2167.5	2167.5	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	-10837.5	-10837.5	0	0	-2167.5	8670	-4335	0	0	0	4335	-2167.5	0	0	0	10837.5	-4335	0	0	-10837.5	-10837.5	-10837.5	0
19	0	0	4335	4335	0	0	8670	4335	17340	0	0	0	-4335	8670	0	0	0	-4335	17340	0	0	4335	4335	4335	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335	0	0	0	-4335	2167.5	0	0	0	-10837.5	4335	0	0	10837.5	10837.5	10837.5	0
23	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335	0	0	0	-4335	2167.5	0	0	0	-10837.5	4335	0	0	10837.5	10837.5	10837.5	0
24	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335	0	0	0	-4335	2167.5	0	0	0	-10837.5	4335	0	0	10837.5	10837.5	10837.5	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

جدول۲ : ماتریس کواریانس ماتریس ***R***

ماتریس ***CovR*** یک ماتریس ۲۵×۲۵ است ولی ماتریس***R*** یک ماتریس ۲۵×۶ است علت این امر این است که ماتریس کواریانس در واقع رابطه بین اعداد درون ستونهای یک ماتریس را بیان می کند (ماتریس ***R*** دارای ۲۵ ستون است.)

با توجه با قاعده ای که در مورد کواریانس دو عدد بیان شد. با توجه با بردارهای جدول ۳ داریم:

3	4	5	6	7	8
255	255	255	255	0	0
255	255	255	255	255	255
0	0	255	255	0	255
255	255	255	255	0	0
255	255	255	255	0	0
255	255	255	255	0	0

جدول ۳ : چند ستون منتخب از ماتریس*R*

به عنوان نمونه عدد ۱۰۸۳۷.۵ واقع در سطر سوم وستون چهارم ماتریس ***CovR***را مورد بررسی قرار می دهیم .مقدار مذکور عددی مثبت است و نشان می دهد اعداد واقع در ستونهای ۳و۴ ماتریس ***R*** به ترتیب دارای تغییرات مشابهی هستند اگربه این دو ستون نگاه کنید متوجه این امر خواهید شد.

نمونه بعدی را عدد ۸۶۷۰- واقع در مختصات (۸و۴) ماتریس***CovR*** در نظر می گیریم.به اعداد واقع در این دو ستون در جدول ۳ توجه کنید نحوه تغییرات اعداد متناظردرسطرهای برابر متناسب با هم نیست و انتظار اینکه کواریانس آنها عددی منفی باشد منطقی خواهد بود.

در مورد عدد ۲۱۶۷.۵ واقع در مختصات (۸و۷) ماتریس***CovR*** نیز به همین شکل است.نکته قابل توجه دیگر اینست که همانطور که بیان شد ماتریس ***CovR***یک ماتریس متقارن نسبت به قطر اصلی است حال با استفاده از فرمولهای ذکر شده چند مقدار ازمقادیر موجود در ماتریس***CovR*** بدست می آوریم اگر اعداد موجود در ستون iماتریس ***CovR*** را با X_i نشان دهیم خواهیم داشت:

$$CovR_{18\times 9} = Cov(X_{18}, X_9) = \frac{1}{5}\sum_{i=1}^{6}(x_{i\times 18} - \overline{X_{18}})(x_{i\times 9} - \overline{X_9})$$

$$\overline{X_{18}} = \frac{1}{6}\sum_{i=1}^{6} R_{i\times 18} = 42.5 \quad , \quad \overline{X_9} = \frac{1}{6}\sum_{i=1}^{6} R_{i\times 9} = 85$$

$$\Rightarrow CovR_{18\times 9} = \frac{1}{5}\sum_{i=1}^{6}(x_{i\times 18} - \overline{X_{18}})(x_{i\times 9} - \overline{X_9}) = \frac{1}{5}[(-42.5\times(-85)) + (-42.5\times 170) + (212.5\times(-85)) + (-42.5\times 170)$$

$$+ (-42.5\times(-85)) + (-42.5\times(-85))] = \frac{1}{5}(3\times(3612.5) - 14450 - 18062.5) = -4335$$

MATLAB Array Editor: CovR
Nov 16, 2007

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335	0	0	0	-4335	2167.5	0	0	0	-10837.5	4335
4	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335	0	0	0	-4335	2167.5	0	0	0	-10837.5	4335
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	2167.5	2167.5	0	0	10837.5	8670	8670	0	0	0	-8670	10837.5	0	0	0	-2167.5	8670
8	0	0	-8670	-8670	0	0	8670	17340	4335	0	0	0	-4335	8670	0	0	0	8670	4335
9	0	0	4335	4335	0	0	8670	4335	17340	0	0	0	-4335	8670	0	0	0	-4335	17340
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

جدول ۴ : قسمتی از ماتریس ***CovR***

$$CovR_{8\times 7} = Cov(X_8 X_7) = \frac{1}{5}\sum_{i=1}^{6}(x_{i\times 8} - \overline{X_8})(x_{i\times 7} - \overline{X_7})$$

$$\overline{X_8} = \frac{1}{6}\sum_{i=1}^{6} R_{i\times 8} = 85, \quad \overline{X_7} = \frac{1}{6}\sum_{i=1}^{6} R_{i\times 7} = 42.5$$

$$\Rightarrow CovR_{8\times 7} = \frac{1}{5}\sum_{i=1}^{6}(x_{i\times 8} - \overline{X_8})(x_{i\times 7} - \overline{X_7}) = \frac{1}{5}[(85\times 42.5) + (170\times 212.5) + (170\times(-42.5)) + (85\times 42.5)$$

$$+(85\times 42.5) + (85\times 42.5)] = \frac{1}{5}(4\times(3612.5) + 36125 - 7225) = 8670$$

MATLAB Array Editor: CovR
Nov 16, 2007

	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335
4	0	0	10837.5	10837.5	0	0	2167.5	-8670	4335
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	2167.5	2167.5	0	0	10837.5	8670	8670
8	0	0	-8670	-8670	0	0	8670	17340	4335
9	0	0	4335	4335	0	0	8670	4335	17340
10	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	0	0	-4335	-4335	0	0	-8670	-4335	-4335
14	0	0	2167.5	2167.5	0	0	10837.5	8670	8670

جدول ۵-۵ : قسمتی از ماتریس ***CovR***

۴-۳. ماتریس بردارهای ویژه PcaCov:

MATLAB Array Editor: PcaCov
Nov 16, 2007

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	-0.3507	-0.127	0.036	-0.0979	0.5554	-0.7223	0.0378	-0.0996	0.0919	-5.6334e-015	-2.2343e-017	2.8354e-017
4	-0.3507	-0.127	0.036	-0.0979	0.1592	0.3457	0.8062	-0.0225	-0.2337	1.815e-014	1.08e-016	-6.0021e-017
5	2.4091e-019	-4.2491e-018	-8.9908e-017	-7.6882e-017	-0.1335	-0.2782	0.0148	0.7914	-0.5275	3.9472e-014	2.1246e-016	-2.2135e-016
6	-2.1052e-021	8.8385e-020	1.318e-017	5.6333e-017	0.0328	0.0615	0.2145	0.5563	0.7998	-5.6869e-014	-2.3741e-016	3.8303e-016
7	-0.1546	0.381	0.1855	-0.3828	-0.3282	-0.161	0.0859	-0.0763	0.0559	-0.7065	0.0026	0.0033
8	0.1961	0.508	0.1495	-0.2848	0.6565	0.322	-0.1719	0.1527	-0.1118	8.5782e-015	2.8059e-017	-4.9719e-017
9	-0.2575	0.381	-0.4735	0.2537	-6.8208e-018	6.4718e-017	1.0106e-016	-8.6601e-018	9.7289e-017	0.0282	0.2365	-0.1186
10	0	0	0	0	4.8541e-023	2.2266e-020	2.6053e-017	2.2712e-017	4.5011e-016	-0.0129	0.915	-0.178
11	0	0	0	0	-4.3056e-026	-1.4721e-022	-8.5665e-019	-1.3299e-018	-8.232e-017	-0.0093	-0.2258	-0.9696
12	0	0	0	0	0	0	0	0	0	1.558e-016	-1.7404e-016	-1.0335e-017
13	0.208	-0.254	-0.6728	-0.663	-6.3395e-017	-6.3351e-017	2.7413e-016	2.5967e-017	7.8954e-017	2.5314e-016	-1.3669e-018	-4.9984e-018
14	-0.1546	0.381	0.1855	-0.3828	-0.3282	-0.161	0.0859	-0.0763	0.0559	0.7065	-0.0026	-0.0033
15	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0
18	0.3507	0.127	-0.036	0.0979	0.0145	-0.1747	0.254	-0.0687	-0.0075	9.6419e-016	1.1682e-017	1.0352e-017
19	-0.2575	0.381	-0.4735	0.2537	-6.8208e-018	6.4739e-017	1.2034e-016	3.5369e-018	-1.0345e-...	-0.0282	-0.2365	0.1186
20	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0
22	-0.3507	-0.127	0.036	-0.0979	-0.0145	0.1747	-0.254	0.0687	0.0075	-9.6419e-016	-1.1682e-017	-1.0352e-017
23	-0.3507	-0.127	0.036	-0.0979	-0.0145	0.1747	-0.254	0.0687	0.0075	-9.6419e-016	-1.1682e-017	-1.0352e-017
24	-0.3507	-0.127	0.036	-0.0979	-0.0145	0.1747	-0.254	0.0687	0.0075	-9.6419e-016	-1.1682e-017	-1.0352e-017
25	0	0	0	0	0	0	0	0	0	0	0	0

	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	0	0	0	0	0	0	0	1	0	0	0	
2	0	0	0	0	0	0	0	0	0	1	0	0	
3	-1.8517e-016	1.5502e-016	0	0	0	0	0	0	0	0	0	0	
4	6.3528e-016	-5.3184e-016	-3.8547e-032	9.2675e-033	3.0078e-034	3.867e-017	0	0	0	0	0	-3.867e-017	
5	1.334e-015	-1.1168e-015	-2.0273e-032	2.5987e-033	8.4342e-035	6.9389e-018	-1.5407e-033	0	0	0	0	-1.3878e-017	
6	-1.9027e-015	1.5929e-015	-5.5311e-032	-7.6449e-033	3.9439e-031	6.9389e-018	1.3096e-032	3.5548e-033	0	0	4.5103e-017	-1.0408e-017	
7	-0.0231	0.0193	-2.5935e-019	3.8138e-020	1.2378e-021	7.5969e-018	9.0237e-035	-6.4907e-034	0	0	3.7266e-018	7.1101e-018	
8	3.0107e-016	-2.5205e-016	3.4532e-033	-1.6372e-033	1.804e-032	-9.8681e-018	-6.872e-034	-5.2779e-034	0	0	-7.921e-019	6.038e-018	
9	-0.5023	0.4206	3.5503e-017	-5.2208e-018	-1.6944e-019	-1.6265e-017	3.6596e-034	4.3476e-033	0	0	1.1211e-017	-9.0538e-018	
10	0.2775	-0.2323	-1.6721e-017	2.4589e-018	7.9803e-020	-2.7756e-017	-6.163e-033	7.2939e-033	0	0	-2.7756e-017	0	
11	0.0718	-0.0601	2.7728e-016	-4.0775e-017	-1.3234e-018	2.7756e-017	6.163e-033	2.516e-033	0	0	1.3878e-017	0	
12	-0.6419	-0.7668	-2.7497e-015	3.3343e-016	1.0821e-017	5.2042e-017	-1.1556e-033	-2.7988e-033	0	0	-6.9389e-018	3.4694e-018	
13	-2.0537e-017	1.7193e-017	3.7483e-033	2.5832e-033	-6.2913e-032	5.0951e-018	3.3807e-034	-1.1352e-033	0	0	-1.3938e-018	1.24e-017	
14	0.0231	-0.0193	2.5935e-019	-3.8138e-020	-1.2378e-021	4.2937e-018	-4.1154e-034	4.7167e-035	0	0	2.3676e-019	-3.2726e-018	
15	1.8173e-015	2.1708e-015	-1	-0.009	-0.0003	-2.7756e-017	0	6.163e-033	0	0	-1.1173e-048	0	
16	0	0	0.0091	-0.9994	-0.0324	-1.6653e-016	-5.5467e-032	-1.2326e-032	0	0	2.2347e-048	0	
17	0	0	-3.425e-016	-0.0324	0.9995	6.6336e-015	1.2757e-030	-2.516e-033	0	0	-1.3878e-017	2.7756e-017	
18	2.8739e-017	-2.406e-017	1.0593e-017	-1.0112e-016	5.6618e-015	-0.866	-1.6653e-016	2.0031e-032	0	0	-6.2546e-018	9.3058e-017	
19	0.5023	-0.4206	-3.5503e-017	5.2208e-018	1.6944e-019	9.3551e-018	-2.3554e-033	-4.2725e-033	0	0	-1.2099e-017	-7.6194e-018	
20	0	0	0	0	0	0	-1	1.282e-016	0	0	-2.3242e-032	5.5511e-017	
21	0	0	0	0	0	0	0	-1	0	0	1.2579e-016	-1.1102e-016	
22	-2.8739e-017	2.406e-017	3.5309e-018	-3.3707e-017	1.8873e-015	-0.2887	-5.5511e-017	2.9606e-016	0	0	0.8165	-2.6818e-016	
23	-2.8739e-017	2.406e-017	3.5309e-018	-3.3707e-017	1.8873e-015	-0.2887	-5.5511e-017	-1.4803e-016	0	0	-0.4082	-0.7071	
24	-2.8739e-017	2.406e-017	3.5309e-018	-3.3707e-017	1.8873e-015	-0.2887	-5.5511e-017	-1.4803e-016	0	0	-0.4082	0.7071	
25	0	0	0	0	0	0	0	0	0	0	0	0	

جدول۶ :ماتریس بردارهای ویژۀ ماتریس ***CovR*** مرتب شده براساس مقادیر ویژه هر بردار

ماتریس ***PcaCov***حاوی بردارهای ویژه استاندارد شده متعامدو مرتب شده براساس مقادیر ویژه است.یعنی ستون اول این ماتریس بردار ویژه متناظر با بزرگترین مقدار ویژه ماتریس ***CovR*** است این ماتریس دارای ۲۵ بردار ویژه است.همانطورکه درجدول ۵-۵ مشاهده می کنید هر چه قدر به ستونهای انتهایی ماتریس ***PcaCov*** نزدیکتر می شویم داده ها کم اهمیتر می شوند چون مقادیر آنها بسیار کوچک شده و بسیار نزدیک به صفرمی شود ،مقادیر آخرین ستون کلاً صفر هستند.حال می توان به جای استفاده از کلیه ستونهای ماتریس فوق از تعداد ستونهای کمتری مثلاً فقط از ۵ ستون اول استفاده کرد. در پروژه ارائه شده در فصل آخر ماتریس بردارهای ویژه مربوط به اعداد ۱۶۰۰×۱۶۰۰ است که ما فقط از ۵۰ ستون اول آن استفاده کرده ایم.

مرحله چهارم انتخاب مؤلفه ها و ساختن *Feature Vector*

در این مرحله مفهوم کاهش ابعاد داده وارد می شود. بردارهای ویژه محاسبه شده که در یک ماتریس۲۵× ۲۵ براساس مقادیر ویژه از بزرگ به کوچک و مؤلفه های داده ها از پر اهمیت به کم اهمیت مرتب شده اند رااگر بخواهیم کاهش بُعد دهیم می توانیم مؤلفه های کم اهمیت را حذف کنیم .
کاری که باید در این مرحله انجام دهیم ایجاد یک ***Feature Vector***است که در واقع ماتریسی از بردارها می باشد این ماتریس شامل آن بردارهای ویژگی است که ما می خواهیم آنها رانگه داریم.
در این مثال ما می خواهیم تنها از ۵ بردار ویژه پر اهمیت ماتریس ویژگیها استفاده کنیم.و آن را در یک ماتریس با نام***FV*** ذخیره می کنیم.

مرحله پنجم بدست آوردن داده های جدید:

در آخرین مرحله فقط باید ترانهاده ***FV*** که در مرحله قبل به دست آوردیم را در ترانهاده داده های نرمال سازی شده ضرب کنیم. البته توجه شود که ماتریس ***NormalR***ترانهاده ماتریس نرمال شده ***R*** است.

MATLAB Array Editor: NormalR
Dec 27, 2007

	1	2	3	4	5	6
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	-42.5	-42.5	212.5	-42.5	-42.5	-42.5
4	-42.5	-42.5	212.5	-42.5	-42.5	-42.5
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	42.5	-212.5	42.5	42.5	42.5	42.5
8	85	-170	-170	85	85	85
9	85	-170	85	-170	85	85
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	-85	170	-85	-85	170	-85
14	42.5	-212.5	42.5	42.5	42.5	42.5
15	0	0	0	0	0	0
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	42.5	42.5	-212.5	42.5	42.5	42.5
19	85	-170	85	-170	85	85
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	-42.5	-42.5	212.5	-42.5	-42.5	-42.5
23	-42.5	-42.5	212.5	-42.5	-42.5	-42.5
24	-42.5	-42.5	212.5	-42.5	-42.5	-42.5

(الف)

MATLAB Array Editor: FV
Dec 27, 2007

	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	0	0
3	-0.3507	-0.127	0.036	-0.0979	0.5554
4	-0.3507	-0.127	0.036	-0.0979	0.1592
5	2.4091e-019	-4.2491e-018	-8.9908e-017	-7.6882e-017	-0.1335
6	-2.1052e-...	8.8385e-020	1.318e-017	5.6333e-017	0.0328
7	-0.1546	0.381	0.1855	-0.3828	-0.3282
8	0.1961	0.508	0.1495	-0.2848	0.6565
9	-0.2575	0.381	-0.4735	0.2537	-6.8208e-018
10	0	0	0	0	4.8541e-023
11	0	0	0	0	-4.3056e-026
12	0	0	0	0	0
13	0.208	-0.254	-0.6728	-0.663	-6.3395e-017
14	-0.1546	0.381	0.1855	-0.3828	-0.3282
15	0	0	0	0	0
16	0	0	0	0	0
17	0	0	0	0	0
18	0.3507	0.127	-0.036	0.0979	0.0145
19	-0.2575	0.381	-0.4735	0.2537	-6.8208e-018
20	0	0	0	0	0
21	0	0	0	0	0
22	-0.3507	-0.127	0.036	-0.0979	-0.0145
23	-0.3507	-0.127	0.036	-0.0979	-0.0145
24	-0.3507	-0.127	0.036	-0.0979	-0.0145
25	0	0	0	0	0

(ب)

جدول۷ :(الف):ماتریس نرمالیزه شدۀ ***R***(ب):پنج بردار ویژه با اهمیتر

$$FV^T \times NormalR \quad Finaldata$$

که ***FV*** ماتریسی است که بردارهای ویـژه در سـطرهای آن بـه ترتیـب مقـادیر ویـژه از مقـادیر زیـاد بـه مقادیرکم قرار گرفته اند و ***NormalR*** ماتریسی است شامل داده هایی که میانگین هر بعـد از عناصـر آن بعدکم شده و ماتریس حاصل ترانهاده شده است .

توجه داشته باشید ماتریس FV^T یک ماتریس ۲۵×۵ است و مـاتریس ***NormalR***یـک مـاتریس ۶×۲۵ است و حاصلضرب دو ماتریس ***Finaldata*** یک ماتریس ۶×۵ خواهد بود.مطابق جدول ۷

MATLAB Array Editor: Finaldata | Page 1
Dec 27, 2007 | 2:29:55 AM

	1	2	3	4	5	6
1	31.5002	244.7002	-555.0548	162.8088	84.5454	31.5002
2	194.3102	-388.6204	-129.5401	-3.7303e-014	129.5401	194.3102
3	-4.0073	-66.8575	12.9803	237.4623	-175.5704	-4.0073
4	67.7143	37.0953	-9.4911	-61.6846	-101.3482	67.7143
5	-2.9088e-014	1.0236e-013	1.6875e-014	-2.5535e-014	-4.5075e-014	-2.9088e-014

جدول۸: ماتریس کاهش بعد یافتهٔ ماتریس ***R***

همانطور که خواهید دید ما در پروژه از ماتریسی مشابه این ماتریس که در پروژه دارای ابعـاد ۵۰×۱۶۰۰ است استفاده خواهیم کرد. با ضرب این ماتریس در ترانهاده ماتریس نرمال سازی شده اولیه ما می توانیم داده های کاهش یافته را برای ورودی شبکه عصبی آماده کنیم.

حال ما با داشتن ماتریس ***FV*** برای نمونه ها می توانیم هر ماتریس دیگر از نمونه های اعـداد صـفر را بـا استفاده از الگوریتم ***PCA*** کاهش بعد دهیم مثلاً اگر ماتریسی مشابه ماتریس ***R1*** با ابعاد ۶×۲۵ داشـته با شیم می توانیم باانجام عملیات زیر داده های پر اهمیت را جدا کرده و تنها از آنها استفاده کنیم.

$$PCAR2_{5\times6} = FV_{25\times5}^{\ T} \times (OrignalMean_{1\times25} - R1_{6\times25})_{6\times25}^{\ T}$$

در این مثال ما با استفاده از الگوریتم ***PCA*** یک ماتریس ۶×۲۵ را بـه یـک مـاتریس ۶×۵ کـاهش بعـد دادیم بدون اینکه داده های پر اهمیت از بین بروند.با استفاده از رابطه زیر مـی تـوانیم مقـادیری کـه از تبدیل ***PCA*** بدست آورده ایم را به داده های اولیه که مقدار میانگین از آنهاکم شده باز گردانیم.

توجه داشته باشید ماتریس FV یک ماتریس ۵×۲۵ است و مـاتریس ***Fianaldata***یـک مـاتریس ۶×۵ است و حاصلضرب دو ماتریس یک ماتریس ۶×۲۵ خواهد بود.وترانهادهٔ آن یک ماتریس ۶×۲۵ که همان ماتریس نرمالیزه شده است.

$$CNormalR_{6\times25} = (FV_{25\times5} \times Finaldata_{5\times6})_{25\times6}^{T}$$

MATLAB Array Editor: CNormalR
Dec 27, 2007

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	0	-42.5	-42.5	-1.7798e-015	2.8244e-015	42.5	85	85	-1.412e-036	1.2524e-039	0	-85	42.5	0	0	0	42.5	85	0	0	-42.5	-42.5	-42.5	0
2	0	0	-42.5	-42.5	-8.7987e-015	4.5323e-015	-212.5	-170	-170	4.9688e-036	-4.4073e-039	0	170	-212.5	0	0	0	42.5	-170	0	0	-42.5	-42.5	-42.5	0
3	0	0	212.5	212.5	-2.2739e-015	1.7985e-016	42.5	-170	85	8.1916e-037	-7.2658e-040	0	-85	42.5	0	0	0	-212.5	85	0	0	212.5	212.5	212.5	0
4	0	0	-42.5	-42.5	-1.3159e-014	-1.1833e-015	42.5	85	-170	-1.2395e-036	1.0994e-039	0	-85	42.5	0	0	0	42.5	-170	0	0	-42.5	-42.5	-42.5	0
5	0	0	-42.5	-42.5	2.9066e-014	-9.491e-015	42.5	85	85	-2.188e-036	1.9407e-039	0	170	42.5	0	0	0	42.5	85	0	0	-42.5	-42.5	-42.5	0
6	0	0	-42.5	-42.5	-1.7798e-015	2.8244e-015	42.5	85	85	-1.412e-036	1.2524e-039	0	-85	42.5	0	0	0	42.5	85	0	0	-42.5	-42.5	-42.5	0

جدول ۹ : داده های بازیابی شده پس از اعمال تکنیک ***PCA***

با اضافه کردن میانگین داده های اولیه را خواهیم داشت:

MATLAB Array Editor: OrignalMean
Dec 27, 2007

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	255	255	212.5	212.5	255	255	42.5	85	85	255	255	0	170	42.5	255	255	0	42.5	85	255	255	212.5	212.5	212.5	255

جدول ۱۰ :میانگین ستونی ماتریس ***R***

$$Rnew_{6\times25} = (FV_{25\times5} \times Finaldata_{5\times6})^{T} + OrignalMean$$

در جدول ۱۰ داده هایی که پس از تبدیل ***PCA*** بازیابی شده اند را مشاهده می کنید همانطور که می بینید ماتریس بدست آمده ***Rnew*** با تقریب بسیار خوب مساوی با ماتریس اولیهٔ ما یعنی ***R*** (جدول ۱۱)است.

MATLAB Array Editor: Rnew
Dec 27, 2007

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	255	255	255	255	255	255	4.9738e-014	7.1054e-014	2.8422e-014	255	255	0	255	4.9738e-014	255	255	0	3.5527e-014	2.8422e-014	255	255	255	255	255	255
2	255	255	255	255	255	255	255	255	255	255	255	0	8.5265e-014	255	255	255	0	2.8422e-014	255	255	255	255	255	255	255
3	255	255	1.1369e-013	2.558e-013	255	255	7.816e-014	255	4.2633e-014	255	255	0	255	7.816e-014	255	255	0	255	4.2633e-014	255	255	2.2737e-013	2.2737e-013	2.2737e-013	255
4	255	255	255	255	255	255	-4.9738e-014	2.8422e-014	255	255	255	0	255	-4.9738e-014	255	255	0	5.6843e-014	255	255	255	255	255	255	255
5	255	255	255	255	255	255	-2.1316e-014	5.6843e-014	4.2633e-014	255	255	0	2.8422e-014	-7.1054e-015	255	255	0	5.6843e-014	4.2633e-014	255	255	255	255	255	255
6	255	255	255	255	255	255	4.9738e-014	7.1054e-014	2.8422e-014	255	255	0	255	4.9738e-014	255	255	0	3.5527e-014	2.8422e-014	255	255	255	255	255	255

جدول۱۱: داد های اولیه پس ازاضافه کردن ماتریس میانگین به داده های بازیابی شده

MATLAB Array Editor: R
Dec 27, 2007

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	255	255	255	255	255	255	0	0	0	255	255	0	255	0	255	255	0	0	0	255	255	255	255	255	255
2	255	255	255	255	255	255	255	255	255	255	255	0	0	255	255	255	0	0	255	255	255	255	255	255	255
3	255	255	0	0	255	255	0	255	0	255	255	0	255	0	255	255	0	255	0	255	255	0	0	0	255
4	255	255	255	255	255	255	0	0	255	255	255	0	255	0	255	255	0	0	255	255	255	255	255	255	255
5	255	255	255	255	255	255	0	0	0	255	255	0	0	0	255	255	0	0	0	255	255	255	255	255	255
6	255	255	255	255	255	255	0	0	0	255	255	0	255	0	255	255	0	0	0	255	255	255	255	255	255

جدول۱۲ : ماتریس ***R***داده های اولیه