Machine Learning

کاهش ابعاد

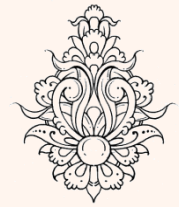
Dimensionality Reduction

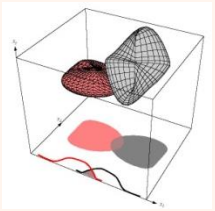
یادگیری ماشین

دانشگاه شهید بهشتی
پژوهشکده‌ی فضای مجازی
پاییز ۱۴۰۲
احمد محمودی ازناوه

فهرست مطالب

- مزایای کاهش ابعاد
- انتخاب خصیصه
- استخراج خصیصه
- تحلیل مؤلفه‌ی اصلی
- تحلیل تفکیک خطی
- روش‌های غیر خطی
- تحلیل عاملی
- تجزیه به مقادیر تکین
- تغییر مقیاس داده‌های چند بعدی





نکبت ابعاد!(مشقت ابعاد)

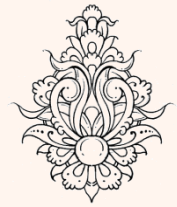
- از لحاظ نظری، افزایش ابعاد منجر به بهبود عملکرد دسته‌بندی می‌شود، اما در عمل همیشه این گونه نیست.

Curse of dimensionality

- انتظار می‌رود در یک فرآیند ایده‌آل دسته‌بندی یا رگرسیون از خصیصه‌های بی‌اهمیت صرفنظر شود و فرآیند «کاهش ابعاد» به صورت جداگانه مورد نیاز نباشد. با این وجود کاهش ابعاد به دلایل زیر مورد توجه قرار می‌گیرد:

مزایای کاهش ابعاد

- «کاهش حجم محاسبات»: حافظه‌ی مصرفی و حجم محاسبات به تعداد (N) و ابعاد (d) داده‌ها بستگی دارد.
  - زمان محاسبات
  - حافظه‌ی مورد نیاز
- «صرفه‌جویی در جمع‌آوری داده»: حذف داده‌های غیرضروری
- «مقاوم‌بودن»(robustness): مدل‌های ساده، هنگامی که داده‌های آموزشی کم باشد، «مقاوم‌تر» می‌باشند؛ قدرت پیش‌بینی برای تعداد مشخصی داده، با افزایش ابعاد، کاهش می‌یابد.

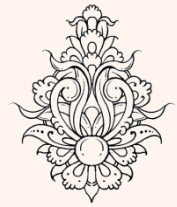


مزایای کاهش ابعاد(ادامه...)

- «استخراج دانش»: با تعداد خصیصه‌های کم‌تر، در مورد داده‌ها و فرآیندهای مربوط به آن درک بهتری وجود خواهد داشت. گاهی این خصیصه‌ها را می‌توان به صورت «عوامل پنهان» در نظر گرفت که متغیرهای قابل مشاهده از آن‌ها نشأت می‌گیرند.

Hidden or latent factor

- هنگامی که تعداد خصیصه‌ها(بدون از دست دادن اطلاعات) کم‌تر باشد، «ساختار داده‌ها» بهتر درک می‌شود. داده‌های پرت و غیرمعمول بهتر تشخیص داده می‌شود؛ قابلیت نمایش بهتری دارند.



(انتخاب –استخراج) خصیصه

Feature Selection vs Extraction

- انتخاب خصیصه:
  - K خصیصه‌ی مهم‌تر (k<d) انتخاب می‌شود.
  - الگوریتم‌های انتخاب زیرمجموعه
- استخراج خصیصه:
  - K خصیصه‌ی جدید، استخراج می‌شود.
  - نگاشت از فضای n-بعدی به فضای k-بعدی
  - روش‌های استخراج خصیصه نیز از دیدگاه‌های مختلفی قابل طبقه‌بندی هستند، روش‌های خطی در برابر روش‌های غیرخطی و یا روش‌های بی‌نظارت در برابر روش‌های بانظارت

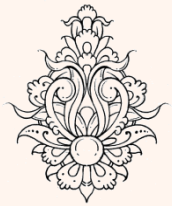
انتخاب زیرمجموعه

- در انتخاب زیرمجموعه، هدف انتخاب **بهترین** زیرمجموعه، زیرمجموعه‌ای با کم‌ترین ابعاد و درست‌ترین نتیجه، می‌باشد.
  - 2^d زیرمجموعه، در یک مجوعه‌ی d-عضوی وجود دارد، بررسی تمام حالات به جز زمانی که d کوچک باشد، امکان‌پذیر نیست.

Forward search

- جستجوی رو به جلو:
  - در گام نخست، مجموعه‌ی خصیصه‌ها، F در حالت اولیه $\emptyset$ در نظر گرفته می‌شود.
  - در هر گام بهترین خصیصه به مجموعه‌ی خصیصه‌ها افزوده می‌شود. (میزان خطای (E(F)) کم‌تر)
  - برای بررسی خطا باید از داده‌های validation استفاده کرد.

$$j = \arg\min_i E(F \cup x_i)$$
$$\text{Add } x_j \text{ to } F \text{ if } E(F \cup x_j) < E(F)$$



انتخاب زیرمجموعه (ادامه...)

- جستجوی رو به عقب: **Backward search**
  - در گام نخست، مجموعه‌ی خصیصه‌ها، F در حالت اولیه تمامی خصیصه‌ها در نظر گرفته می‌شود.
  - در هر گام بدترین خصیصه از مجموعه‌ی خصیصه‌ها حذف می‌شود.
- هنگامی که تعداد خصیصه‌ها زیاد است، روش جستجوی رو به جلو ترجیح داده می‌شود.
- انتخاب زیرمجموعه به صورت بانظارت است.
- در کاربردهایی که یک خصیصه به تنهایی اطلاعات مفیدی ندارد، انتخاب خصیصه مفید نیست.(مانند تشخیص چهره)

Iris data: Single feature

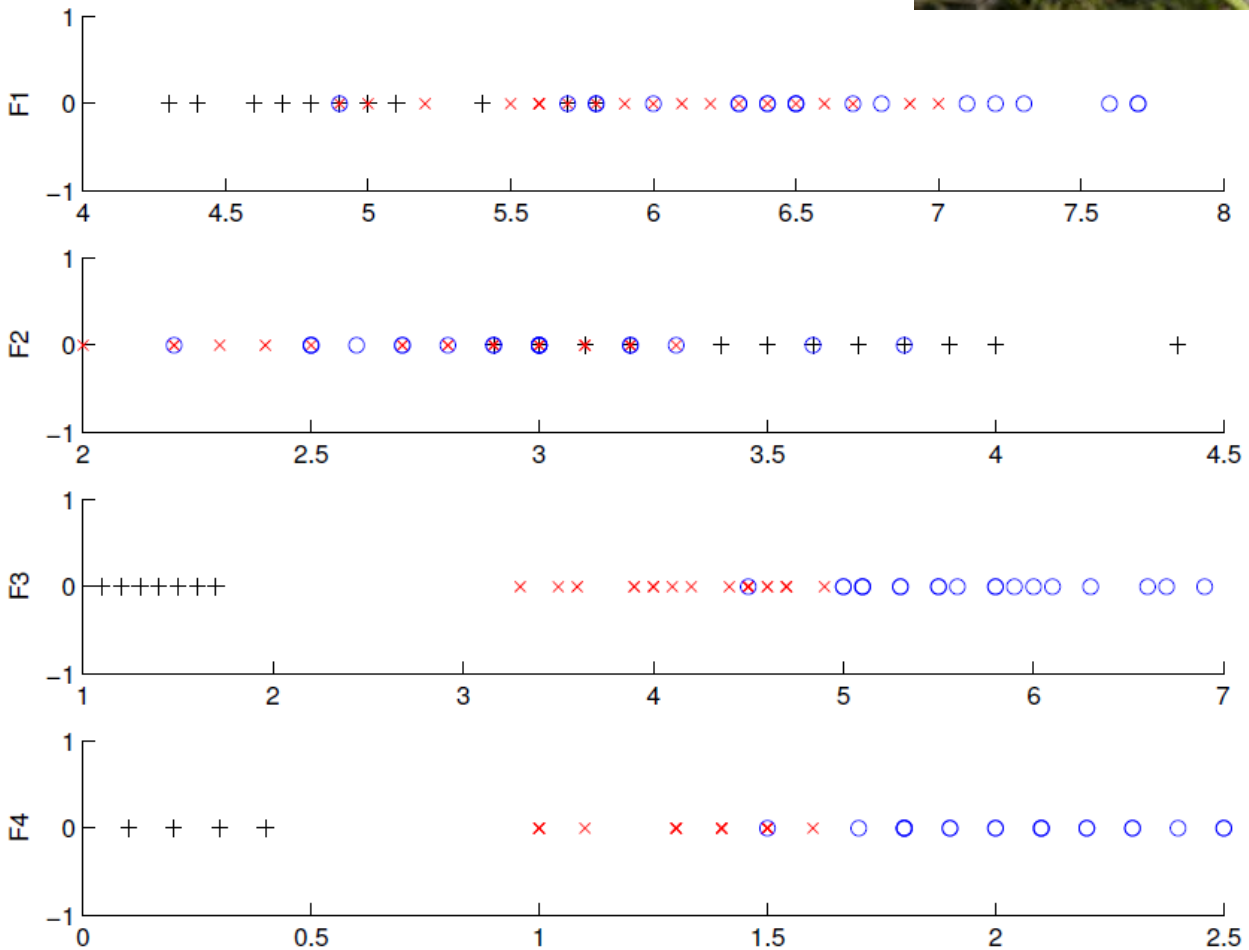
مثال

0.76

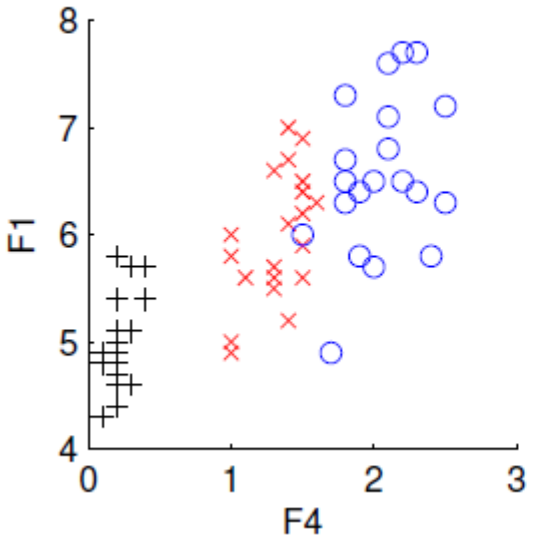
0.57

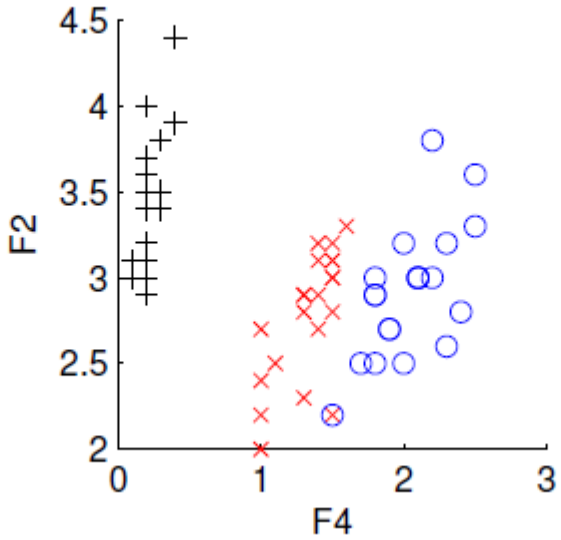
0.92

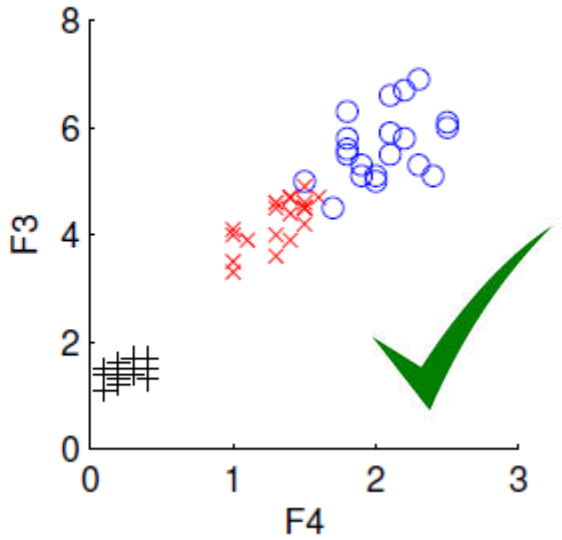
0.94



Iris data: Add one more feature to F4







0.87 ***0.94*** ***0.96***

در صورت اضافه کردن خصیصه‌ی بعدی نتایج افت می‌کند!

در بسیاری موارد انتخاب خصیصه‌ها به نوع دسته‌بند بستگی دارد.

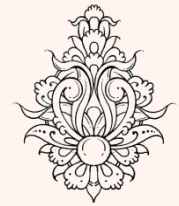
در صورت کوچک بودن دسته‌داده، خصیصه‌ی انتخاب شده، می‌تواند به نحوه‌ی تقسیم پایگاه به دو دسته‌ی training و validation مربوط باشد.

تحلیل مؤلفه‌های اصلی

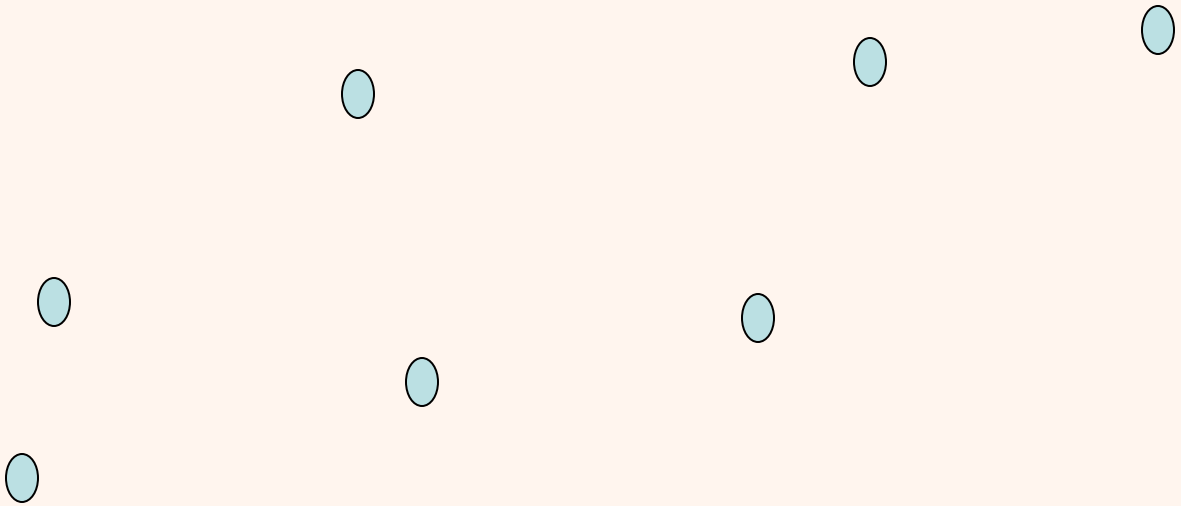
- هدف نگاشت داده‌ی d-بعدی به فضای k-بعدی است (k<d)، به گونه‌ای که کم‌ترین میزان اتلاف رخ دهد.
  - نگاشت x در راستای w:

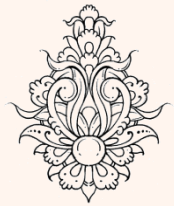
$$z = w^T x$$

- این راستا به گونه‌ای انتخاب می‌شود که Var(z) بیشینه شود، راستایی که داده در امتداد آن بیشترین تغییرات را داشته باشد.
  - این مسأله باعث می‌شود، تفاوت نمونه‌ها آشکارتر شود.
- این شیوه‌ی کاهش بعد به صورت «بی‌نظارت» است.



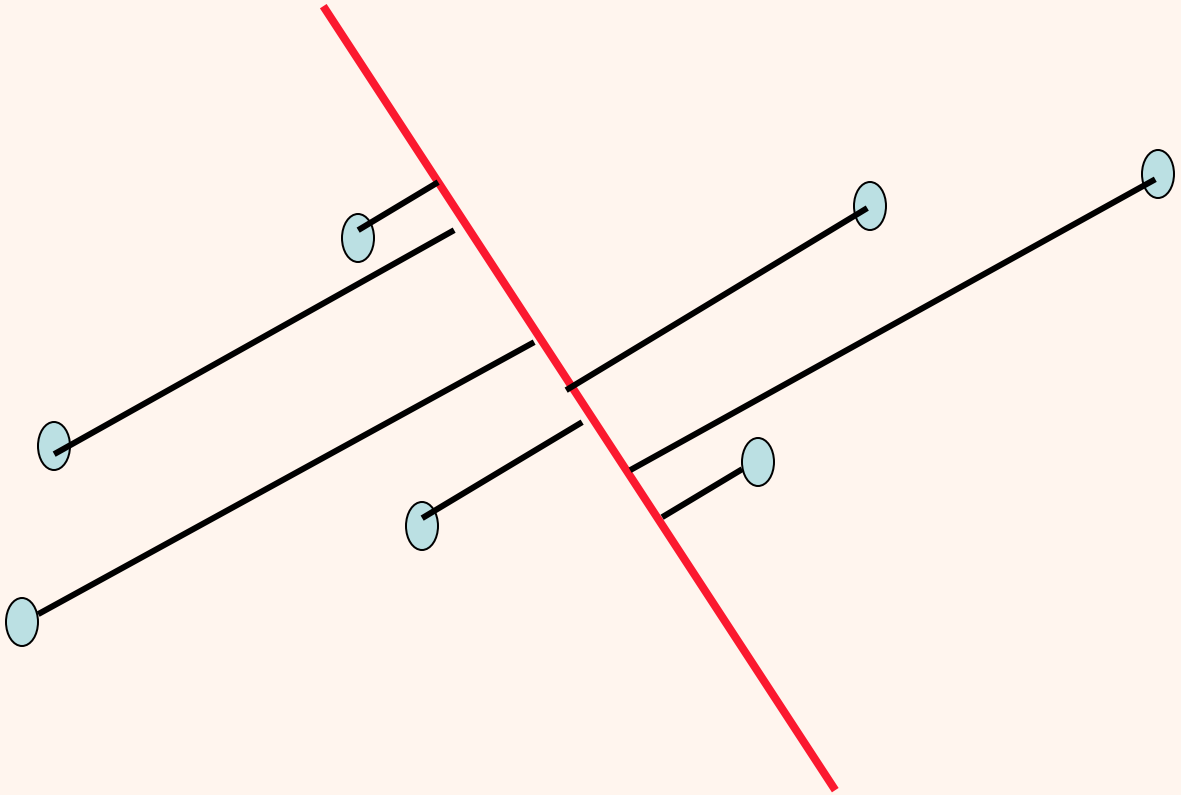
تحلیل مؤلفه‌های اصلی(ادامه...)





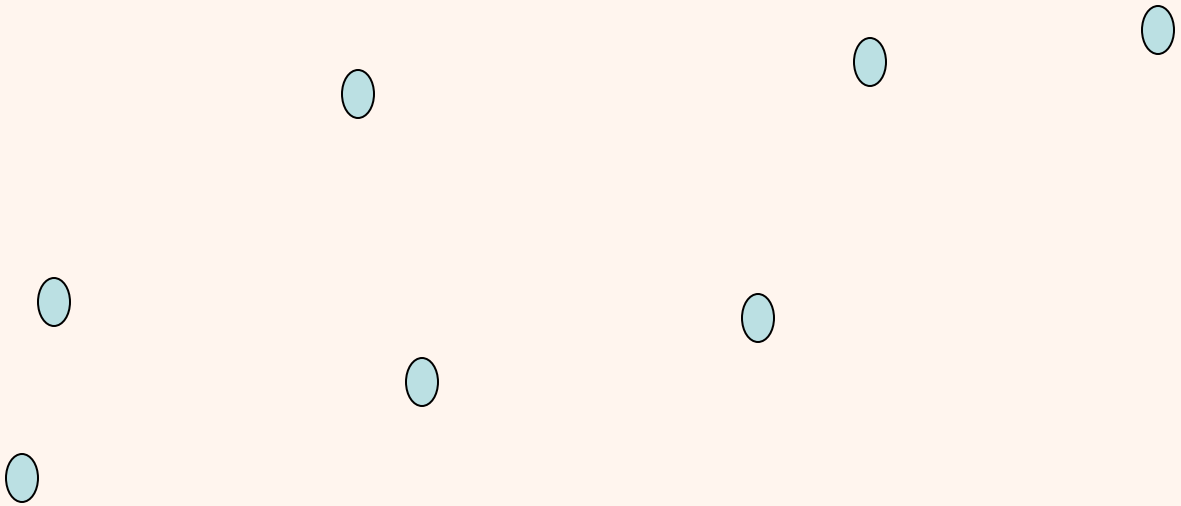
http://www.public.asu.edu/~jye02/

تحلیل مؤلفه‌های اصلی(ادامه...)



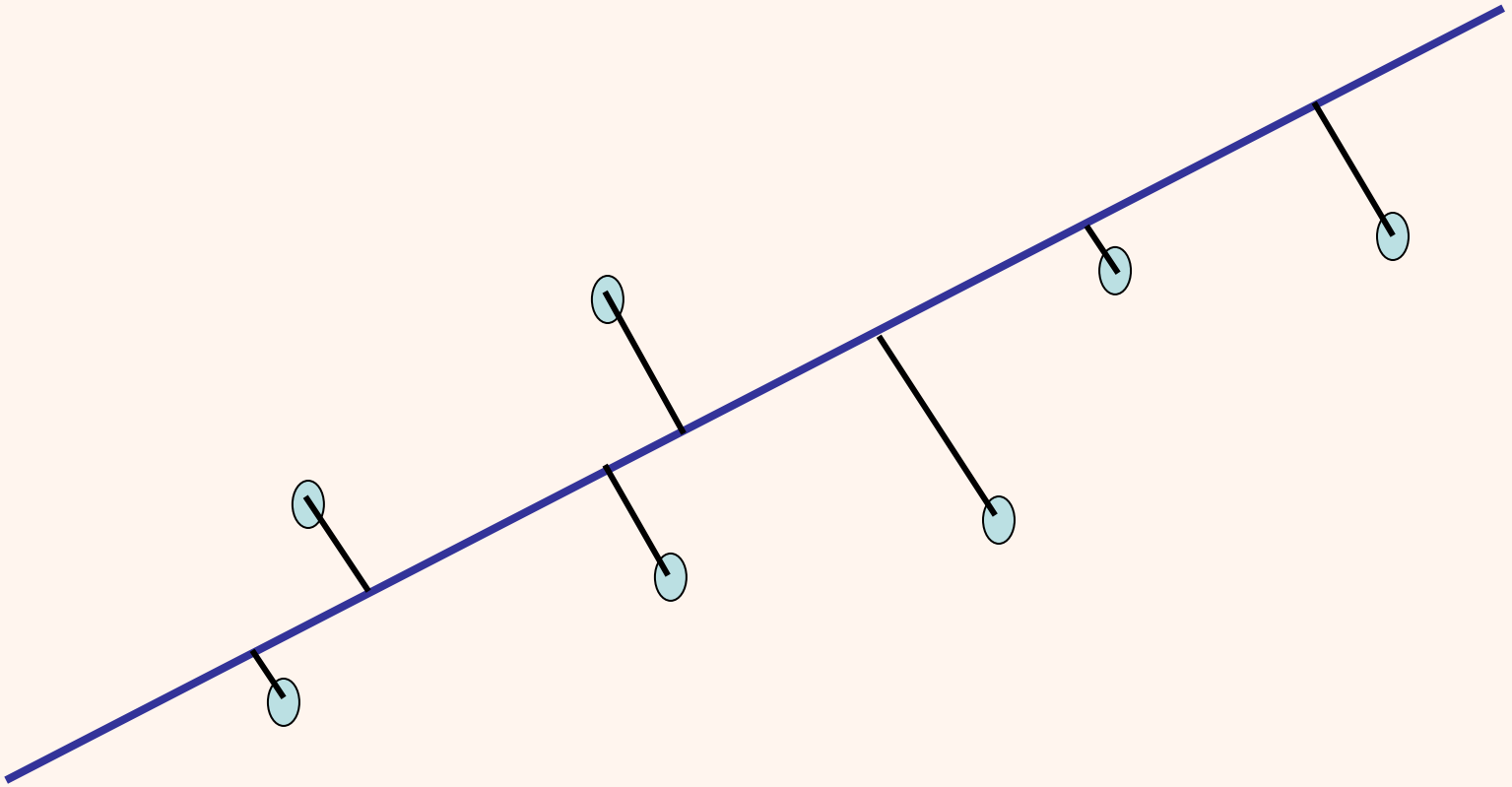
http://www.public.asu.edu/~jye02/

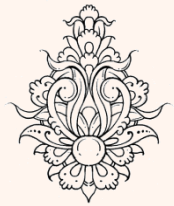
تحلیل مؤلفه‌های اصلی(ادامه...)



http://www.public.asu.edu/~jye02/

تحلیل مؤلفه‌های اصلی(ادامه...)





http://www.public.asu.edu/~jye02/

تحلیل مؤلفه‌های اصلی(ادامه...)

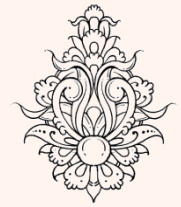
- در راستای w، پراکندگی داده ماکزیمم می‌شود:

$$\begin{aligned}\mathrm{Var}(z) = \mathrm{Var}(w^T x) &= \mathrm{E}[(w^T x - w^T \mu)^2] \\ &= \mathrm{E}[(w^T x - w^T \mu)(w^T x - w^T \mu)] \\ &= \mathrm{E}[w^T (x - \mu)(x - \mu)^T w] \\ &= w^T \mathrm{E}[(x - \mu)(x - \mu)^T] w \\ &= w^T \Sigma w\end{aligned}$$

where $\mathrm{Cov}(x) = \Sigma$

- در این حالت تنها راستا است که اهمیت دارد، در نتیجه برای یافتن پاسخ یکتا، باید شرط زیر نیز برقرار باشد:

$$\|w\| = 1$$



تحلیل مؤلفه‌های اصلی(ادامه...)

- در نتیجه برای اولین مؤلفه‌ی اساسی رابطه‌ی زیر به دست می‌آید: $\max_{\mathbf{w}_1} \mathbf{w}_1^T \Sigma \mathbf{w}_1 - \alpha\left(\mathbf{w}_1^T \mathbf{w}_1 - 1\right)$
- با مشتق گرفتن نسبت به $\boldsymbol{w_1}$ و برابر صفر قرار دادن آن

$$2\Sigma \mathbf{w}_1 - 2\alpha \mathbf{w}_1 = 0$$

- در نتیجه

$$\Sigma \mathbf{w}_1 = \alpha \mathbf{w}_1$$

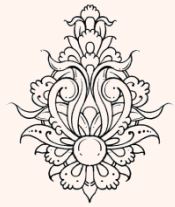
در نتیجه W_1 یکی از بردارهای ویژه‌ی ماتریس Σ می‌باشد

- از طرفی $\mathbf{w}_1^T \Sigma \mathbf{w}_1 = \alpha$ ، در واقع واریانس در راستای $\mathbf{w}_1$ برابر مقدار ویژه‌ی متناظر با آن است.

اولین مؤلفه‌ی اصلی، برابر بردار ویژه‌ی ماتریس کواریانس با بیشترین مقدار ویژه است.

- برای یافتن دومین مؤلفه‌ی اصلی، علاوه بر شرایط پیش باید بر راستای اولین مؤلفه‌ی اساسی هم عمود باشد، در این حالت داده‌های نگاشت شده «ناهمبسته» (uncorrelated) خواهند بود.

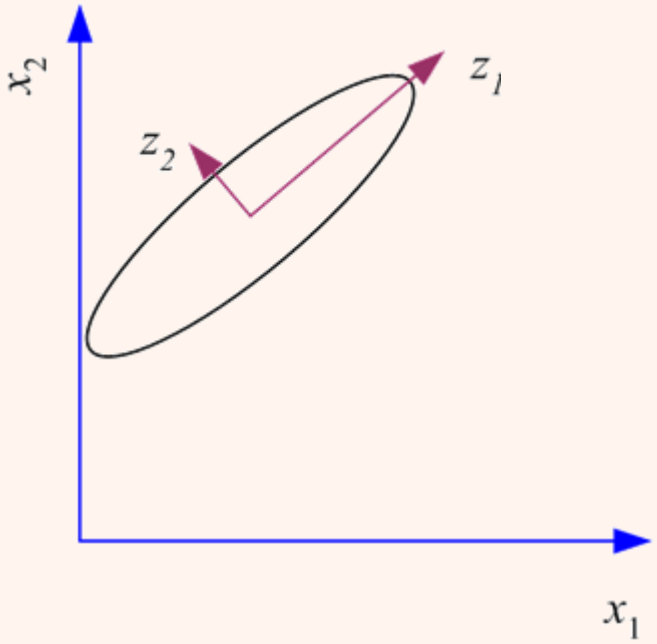
دومین مؤلفه‌ی اصلی، برابر بردار ویژه‌ی ماتریس کواریانس با بیشترین مقدار ویژه در رده‌ی دوم است، به همین ترتیب سایر مقادیر ویژه به دست می‌آیند.



تحلیل مؤلفه‌های اصلی(ادامه...)

$$z = W^T(x - m)$$

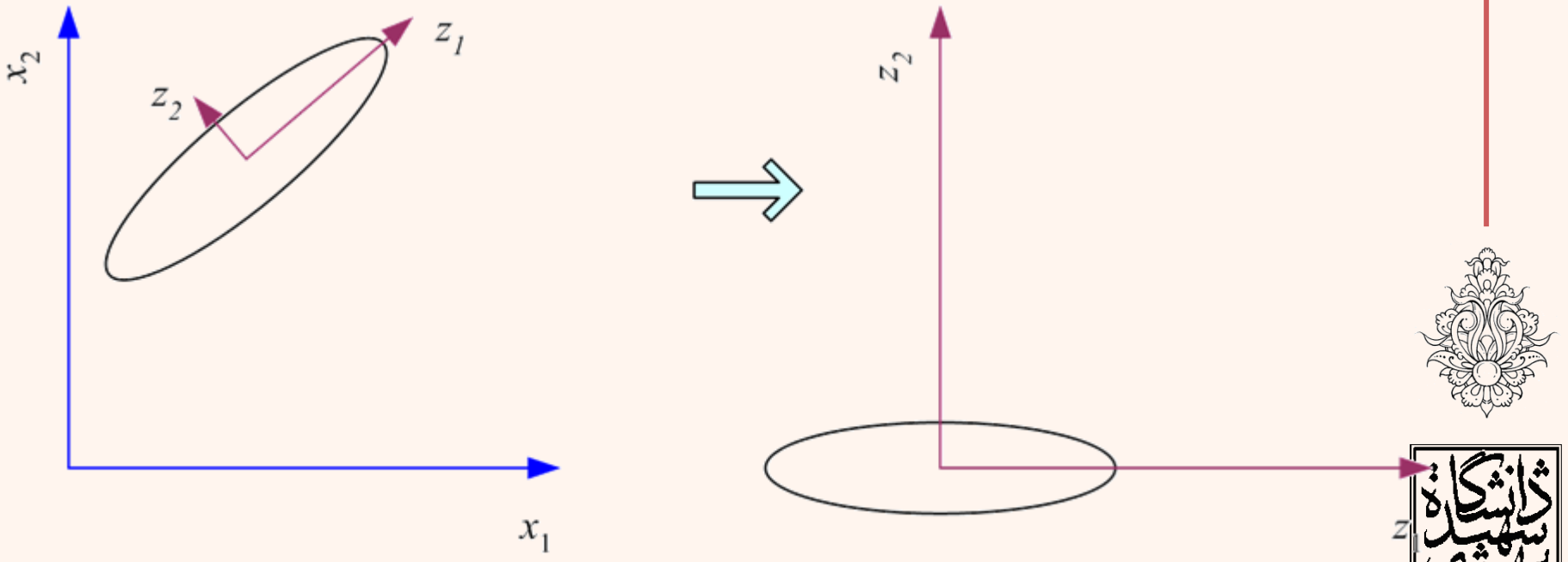
- ستون‌های W، بردارهای ویژه‌ی ماتریس کواریانس هستند.



تحلیل مؤلفه‌های اصلی(ادامه...)

$$z = W^T(x - m)$$

- ستون‌های W، بردارهای ویژه‌ی ماتریس کواریانس هستند.



کاهش بعد

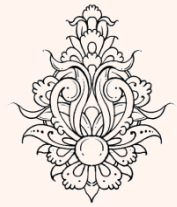
- در صورتی برخی مقادیر ویژه، کوچک باشند. داده‌ها در راستای بردار ویژه‌ی متناظر با آن واریانس کمی دارد و قابل صرفنظر کردن است.
  - در این حالت K مؤلفه‌ی پرارزش انتخاب می‌شوند، با فرض آن که مقادیر ویژه به صورت صعودی مرتب شده باشند.

Proportion of Variance (PoV)

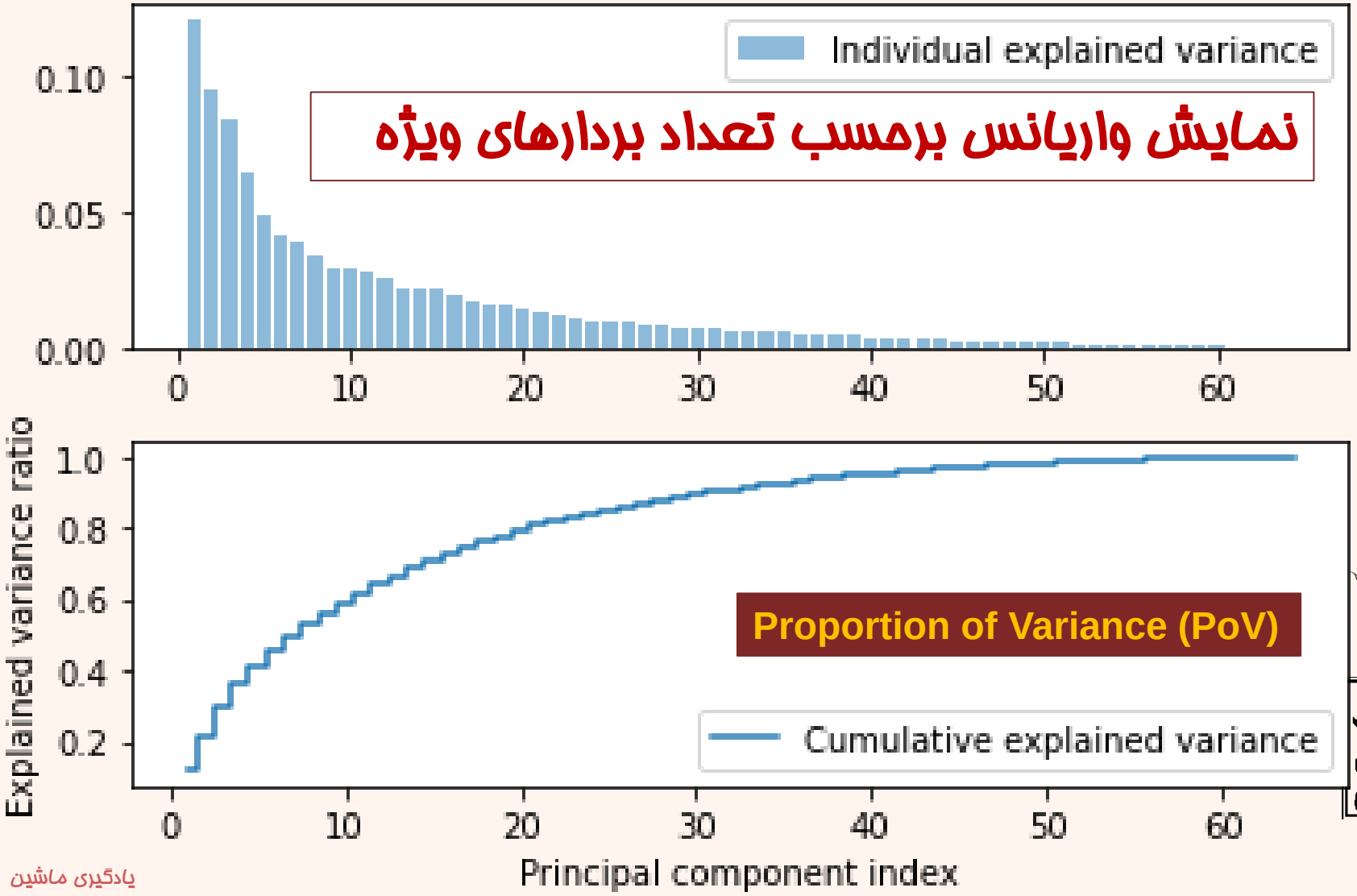
$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_k + \cdots + \lambda_d}$$

PoV>0.9

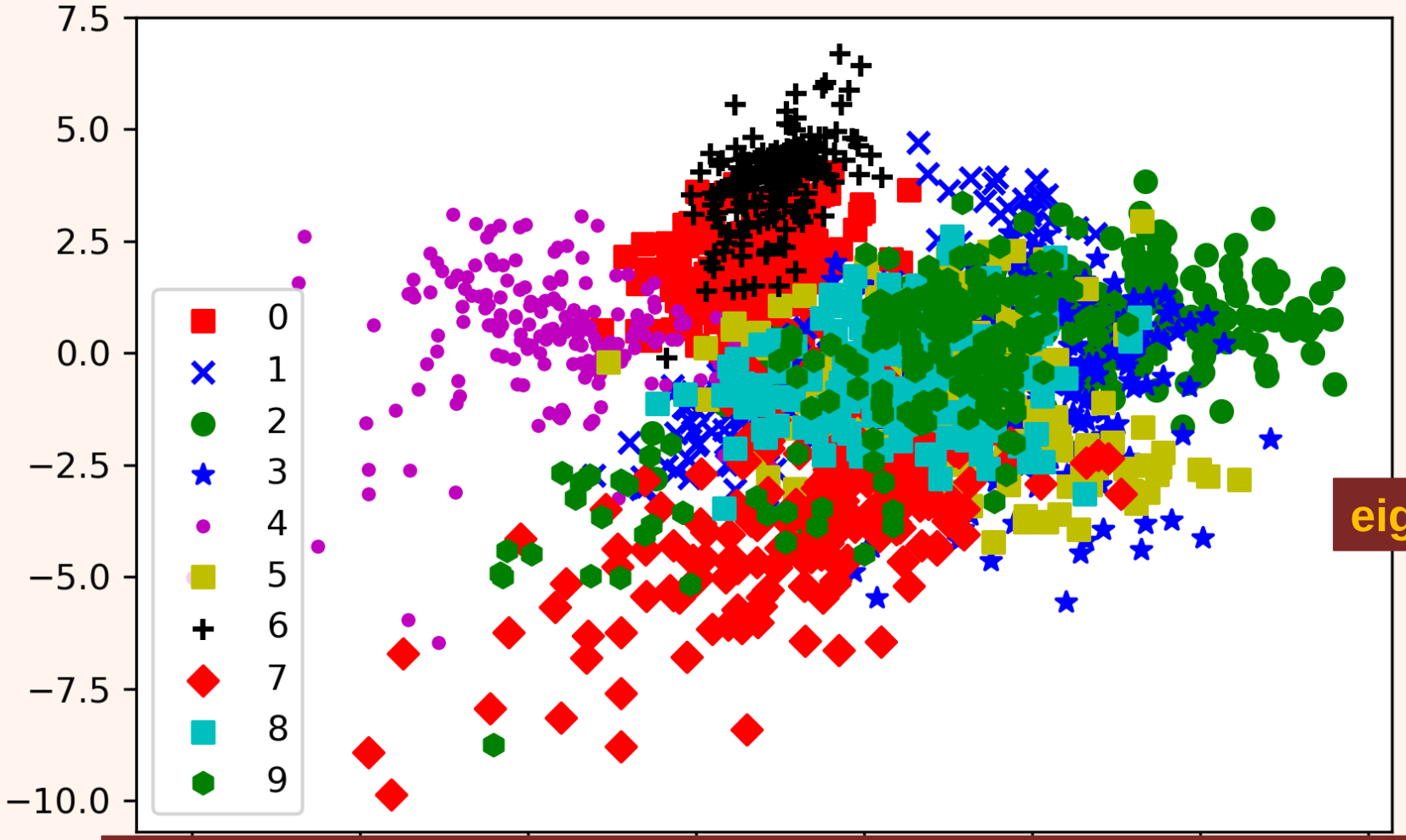
  - در کاربردهای نظیر پردازش تصویر یا صوت، معمولاً کاهش ابعاد قابل توجه است.



Scree graph


Individual explained variance
نمایش واریانس برحسب تعداد بردارهای ویژه
0.10
0.05
0.00
0
10
20
30
40
50
60
Explained variance ratio
1.0
0.8
0.6
0.4
0.2
Proportion of Variance (PoV)
Cumulative explained variance
Principal component index

Visual analysis

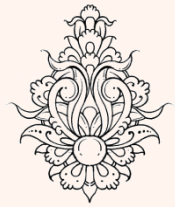


eigendigits

در صورتی که سه بعد نخست، حاوی بخش عمده‌ای از واریانس باشند،
می‌توان داده‌ها از آن‌ها برای «بررسی دیداری» بهره برد.

چند نکته

- علاوه بر در نظر گرفتن PoV، می‌توان بردارهای ویژه‌ای که مقدار ویژه‌ی متناظر آن از یک حدآستانه (به عنوان مثال میانگین واریانس) کم‌تر است را حذف نمود.
- در صورتی که واریانس در ابعاد مختلف تغییرات زیادی داشته باشند، بیش از مقدار همبستگی بر روی مؤلفه‌ی اصلی اثرگذار خواهد بود.
  - در این شرایط می‌توان از بردارها و مقادیر ویژه‌ی «ماتریس همبستگی» (R) استفاده کرد یا این که داده‌ها را به گونه‌ای نرمال کرد که همگی واریانس یکسان داشته باشند.

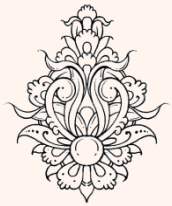


چند نکته

- PCA، نسبت به نویز به شدت حساس است.
  - یک روش ساده حذف داده‌های پرت با استفاده از فاصله‌ی Mahalanobis پیش از محاسبه‌ی ماتریس کواریانس است.
- در میان تمام بردارهای متعامد، PCA کم‌ترین میزان خطا را دارد.

Reconstruction error $\sum_t \left\| \hat{\mathbf{x}}^t - \mathbf{x}^t \right\|$

- Karhunen-Loève expansion و Hotelling transform نام‌های دیگری هستند برای مفاهیم مشابه به کار می‌روند.



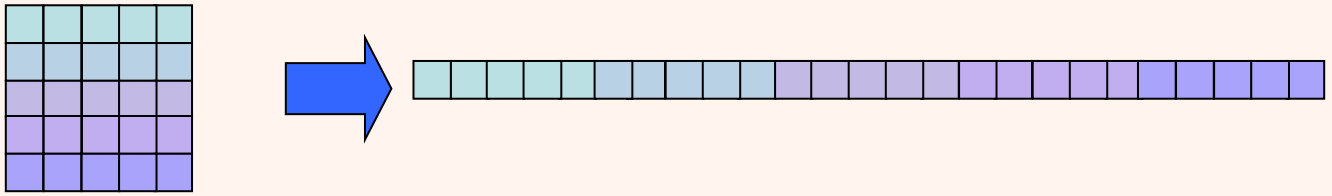
کاربرد PCA در شناسایی چهره

پایگاه داده‌ی ORL

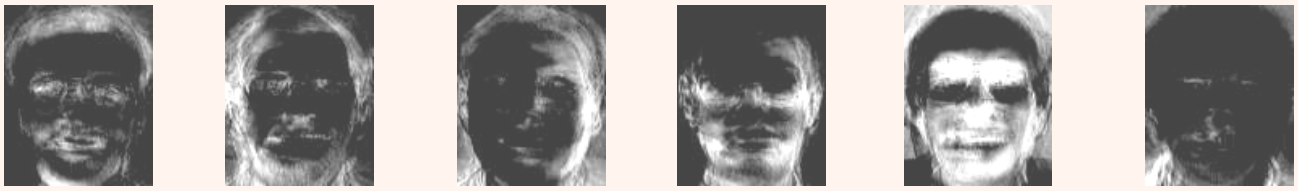
میانگین چهره‌ها

M. Turk, A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

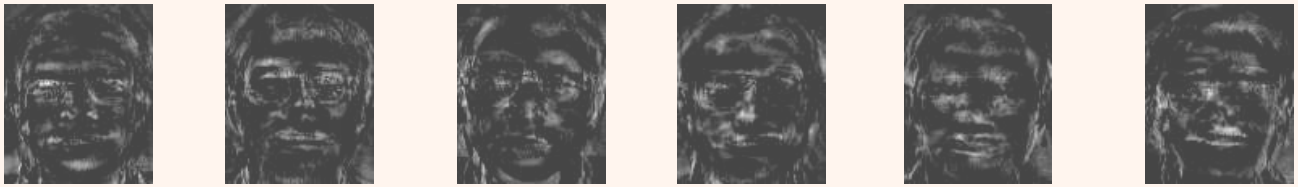
کاربرد PCA در شناسایی چهره(ادامه...)



Eigenfaces



۱ ۲ ۳ ۴ ۵ ۶



۳۵ ۳۶ ۳۷ ۳۸ ۳۹ ۴۰

کاربرد PCA در شناسایی چهره(ادامه...)

تصویراصلی

=

تصویرمیانگین

+(2.4

اولین مؤلفه‌ی اساسی

−0.05

دومین مؤلفه‌ی اساسی

+1.9

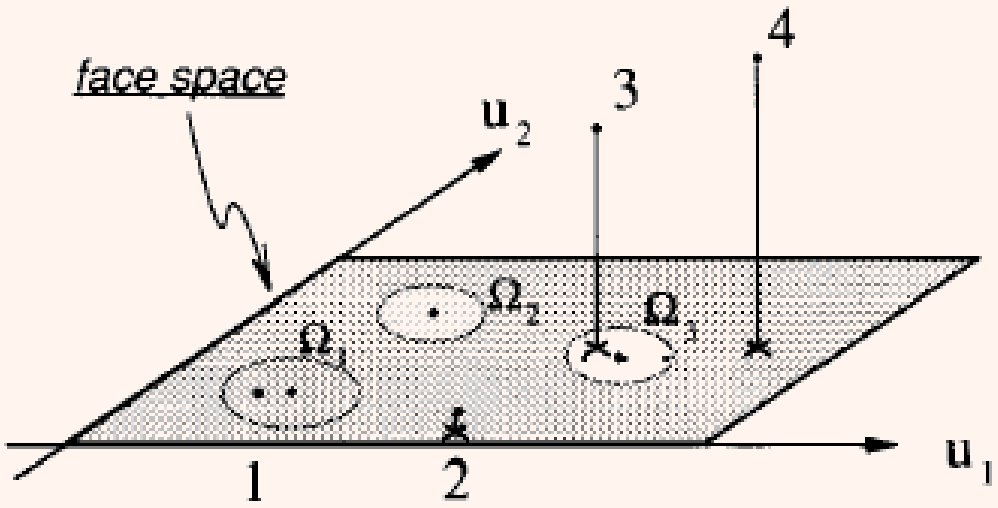
سومین مؤلفه‌ی اساسی

+ .... −0.012

)10^3

آخرین مؤلفه‌ی اساسی

تشخیص چهره



$$\|\hat{\mathbf{x}} - \mathbf{x}\|$$