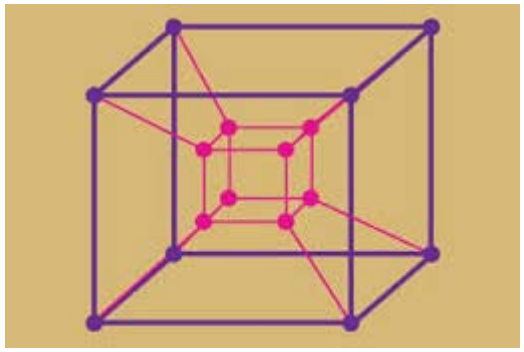
کاهش ابعاد
Dimension Reduction

عادل قاضی خانی
هیئت علمی گروه کامپیوتر دانشگاه امام رضا (ع)

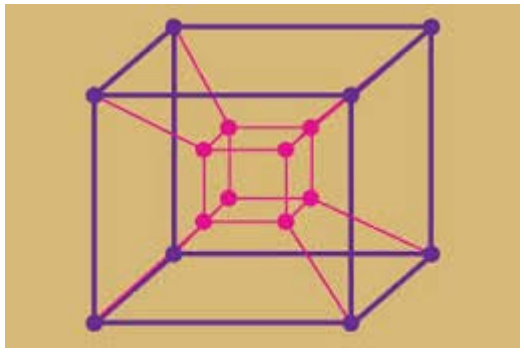


کاهش ابعاد

- تغییر در ابعاد داده های ورودی
- داده های حجیم (داده های متنی، DNA و ....)

- **چرا؟**

1. feature selection
2. feature extraction



کاهش ابعاد - دلایل

- کاهش پیچیدگی الگوریتم یادگیری
- هزینه استخراج
- مدلهای ساده تر بر روی داده های کوچکتر مقاوم تر هستند
- یادگیری بهتر
- تحلیل بهتر داده ها

1. feature selection
2. feature extraction

کاهش ابعاد

- **دو شکل اصلی:**

- انتخاب ویژگی[1]
- استخراج ویژگی[2]

1. feature selection
2. feature extraction

انتخاب ویژگی

- اگر ابعاد داده های ورودی d باشد، در انتخاب ویژگی می خواهیم k بعد را که بیشترین اطلاعات را به ما می دهد انتخاب کنیم و $(d-k)$ بعد را حذف کنیم

استخراج ویژگی

- اگر ابعاد داده های ورودی d باشد، در استخراج ویژگی می خواهیم k بعد جدید را با استفاده از d بعد اولیه ایجاد کنیم

انتخاب ویژگی

- **روش‌ها**

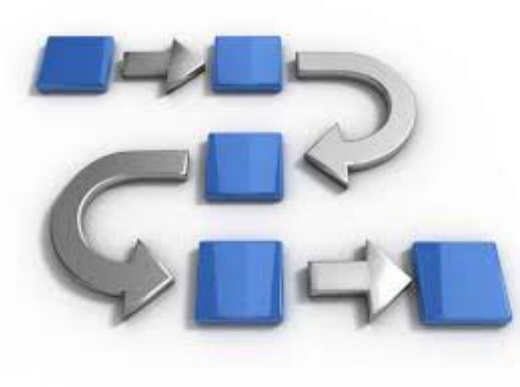
❖فیلتر

❖پوشش

❖ترکیب فیلتر و پوشش

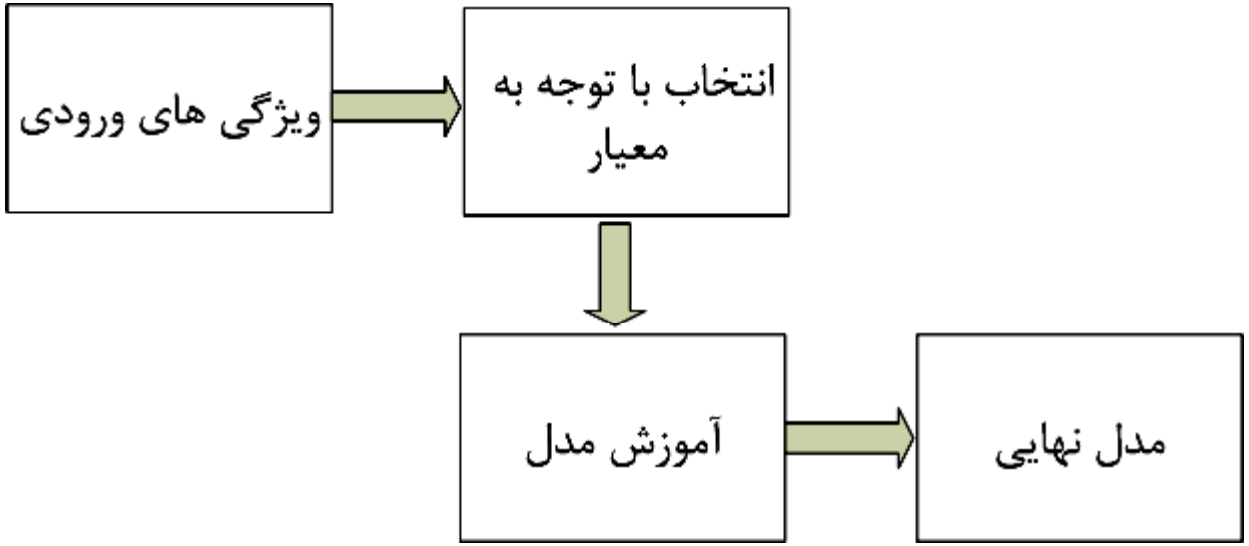
❖تعبیه شده

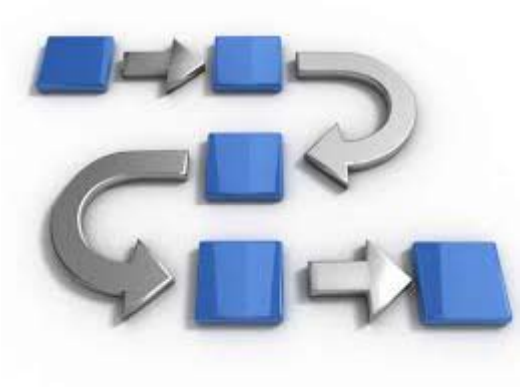
1-Filter
2-Wrapper
3-Embedded



انتخاب ویژگی – روش فیلتر

- در این روش‌ها با استفاده از معیارهایی رتبه بندی ویژگی ها انجام شده و سپس انتخاب ویژگی انجام می شود

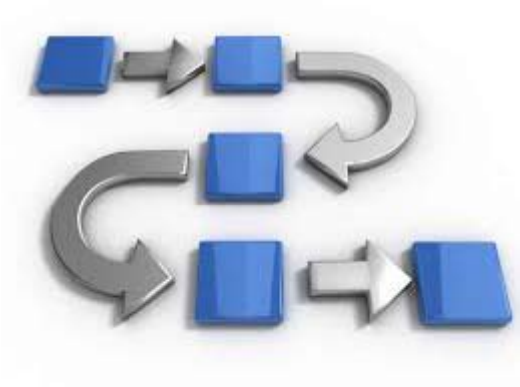




انتخاب ویژگی – روش فیلتر

- معیار پیرسون برای ویژگی i ام

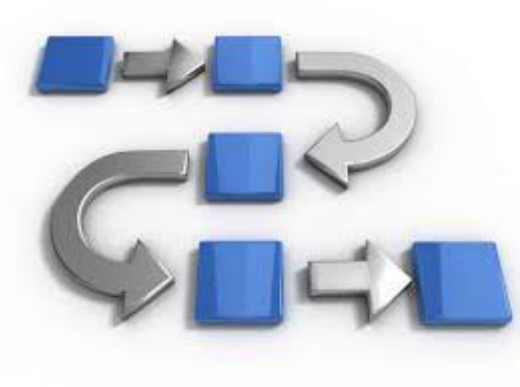
$$R(i) = \frac{cov(x_i, Y)}{\sqrt{var(x_i) * var(Y)}}$$



انتخاب ویژگی – روش فیلتر

- **انواع معیارها**

۱. مبتنی بر فاصله
۲. مبتنی بر اطلاعات و عدم قطعیت
۳. مبتنی بر احتمال و خطا
۴. مبتنی بر وابستگی
۵. ...



انتخاب ویژگی – روش فیلتر

- **معیارهای مبتنی بر فاصله**
- ویژگی ارجحیت دارد که مقدار فاصله ای که بین کلاسهای قرار می دهد بیشتر باشد

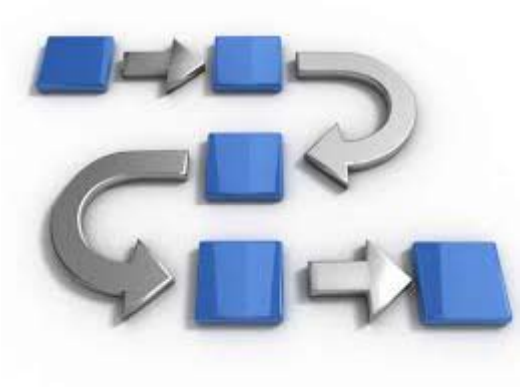
Kullback-Liebler : $D_{KL}(p,q) = \int p(x)\log_2\left(\frac{p(x)}{q(x)}\right)dx$

Bhattacharya : $D_B(p,q) = \int \sqrt{p(x) \times q(x)}\,dx$

Jeffrey'sdivergence: $D_J(p,q) = \int (p(x) - q(x))\left(\log_2\left(\frac{p(x)}{q(x)}\right)\right)dx$

Matusita: $D_M(p,q) = \int \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx$

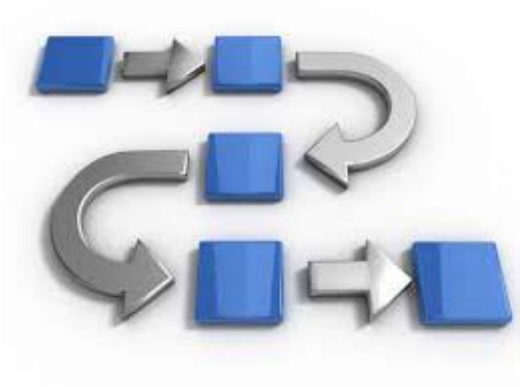
Kagan's divergence: $D_K(p,q) = \frac{1}{2}\int \frac{(p(x)-q(x))^2}{p(x)}dx$



انتخاب ویژگی – روش فیلتر

- **معیارهای مبتنی بر اطلاعات و عدم قطعیت**
- ویژگی که ارزش اطلاعات بیشتری داشته باشد.
- معیار آنتروپی

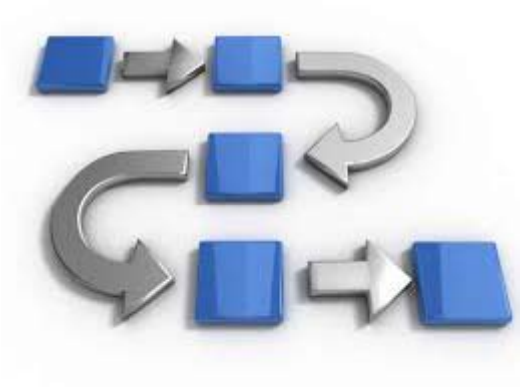
$$H(A) = -\sum_{i=1}^{n} p_i \log_2 p_i$$



انتخاب ویژگی – روش فیلتر

- **معیارهای مبتنی بر احتمال خطا**
- مجموعه ویژگی هایی که احتمال خطا در آنها کمتر باشد
- مثلاً

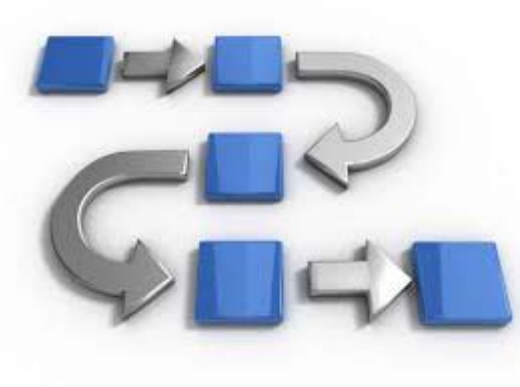
$$P(error) = \int_{\mathcal{R}_2} p(x|w_1)p(w_1)\,dx + \int_{\mathcal{R}_1} p(x|w_2)p(w_2)\,dx$$



انتخاب ویژگی – روش فیلتر

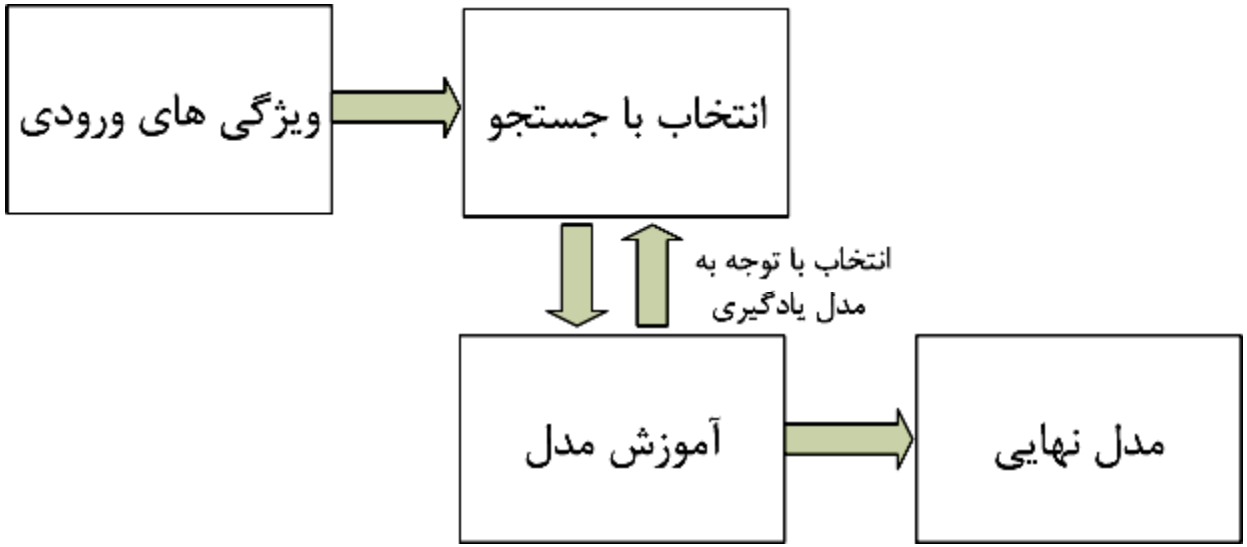
- **معیارهای مبتنی بر وابستگی**
- ویژگی هایی که با کلاسها وابستگی بیشتر دارند ارجحیت دارند
- مثلاً: معیار پیرسون

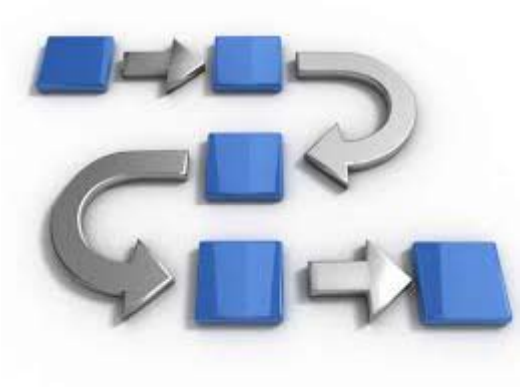
$$R(i) = \frac{cov(x_i, Y)}{\sqrt{var(x_i) * var(Y)}}$$



انتخاب ویژگی – روش پوشش

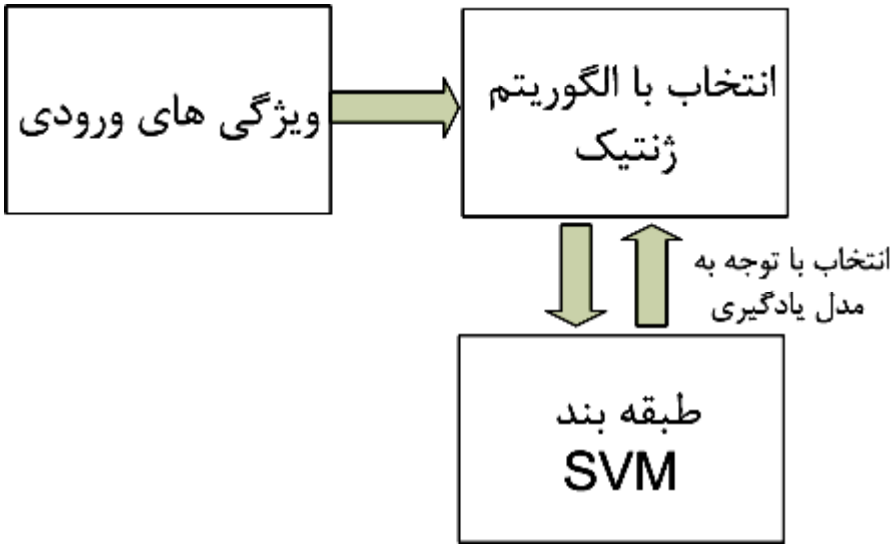
- از مدل یادگیری برای تایید ویژگی های انتخاب شده استفاده می کند





انتخاب ویژگی – روش پوشش

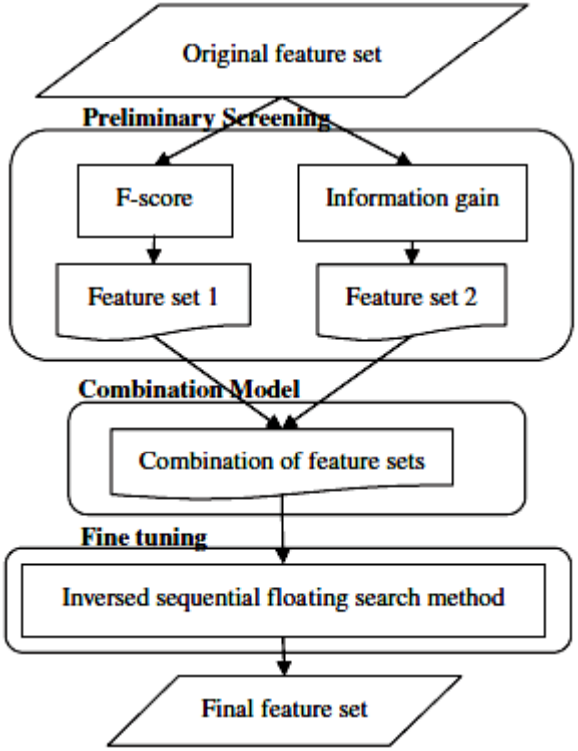
- از مدل یادگیری برای تایید ویژگی های انتخاب شده استفاده می کند



انتخاب ویژگی – ترکیب فیلتر و پوشش

- از بتدا با یک روش فیلتر بهترین ویژگی ها انتخاب می شوند سپس از یک روش پوشش استفاده می شود تا ویژگی های بهتر گزینش شوند

انتخاب ویژگی – ترکیب فیلتر و پوشش



Hui-Huang Hsu, Cheng-Wei Hsieh, Ming-Da Lu, Hybrid feature selection by combining filters and wrappers, Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8144-8150

انتخاب ویژگی – ترکیب فیلتر و پوشش

- **بخش فیلتر**

- معیار F-score برای تعیین قدرت تفکیک ویژگی

$$F(i) \equiv \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_1 - 1}\sum_{k=1}^{n-}(x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2}$$

Hui-Huang Hsu, Cheng-Wei Hsieh, Ming-Da Lu, Hybrid feature selection by combining filters and wrappers, Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8144-8150

انتخاب ویژگی – ترکیب فیلتر و پوشش

- **بخش فیلتر–ادامه**

- معیار ارزش اطلاعات برای تعیین ارزش اطلاعاتی ویژگی

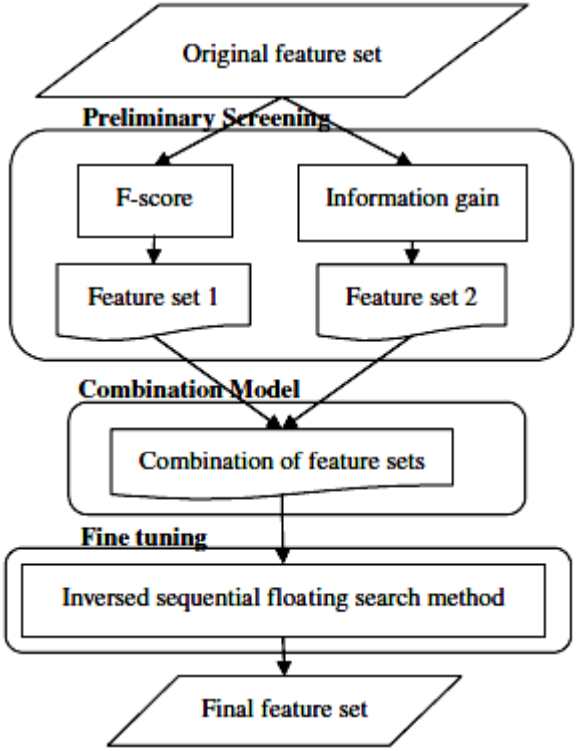
$$\text{Entropy}(N) = \sum_{i=1}^{k} P_i \log_k \left(\frac{1}{P_i}\right) = -\sum_{i=1}^{k} P_i \log_k P_i$$

$$\text{Entropy}(D_j) = \sum_{i=1}^{|D_j|} \frac{D_{ji}}{N} \times \text{Entropy}(D_{ji})$$

$$\text{IG}(D)_j = \text{Entropy}(N) - \text{Entropy}(D)$$

Hui-Huang Hsu, Cheng-Wei Hsieh, Ming-Da Lu, Hybrid feature selection by combining filters and wrappers, Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8144-8150

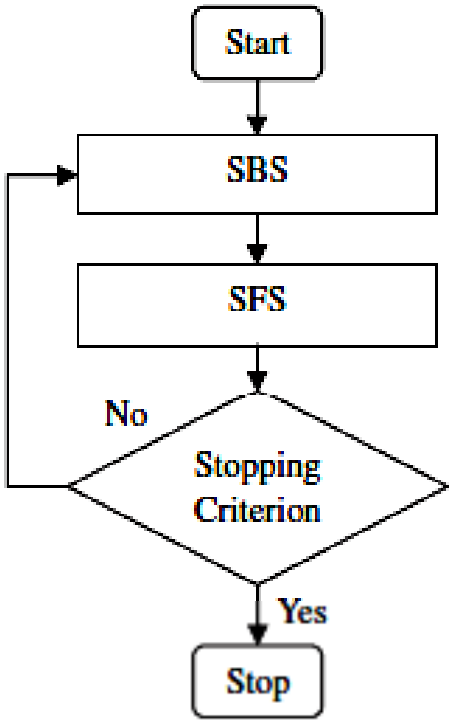
انتخاب ویژگی – ترکیب فیلتر و پوشش



Hui-Huang Hsu, Cheng-Wei Hsieh, Ming-Da Lu, Hybrid feature selection by combining filters and wrappers, Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8144-8150

انتخاب ویژگی – ترکیب فیلتر و پوشش

- **بخش پوشش**



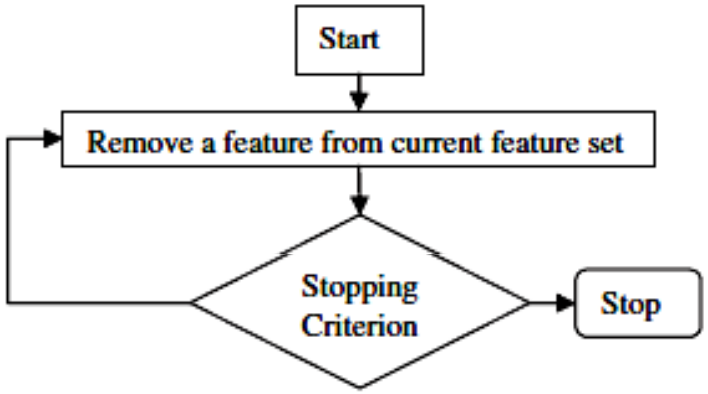
- شرط پایان بر اساس میزان خطا در مجموعه تصدیق

Hui-Huang Hsu, Cheng-Wei Hsieh, Ming-Da Lu, Hybrid feature selection by combining filters and wrappers, Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8144-8150

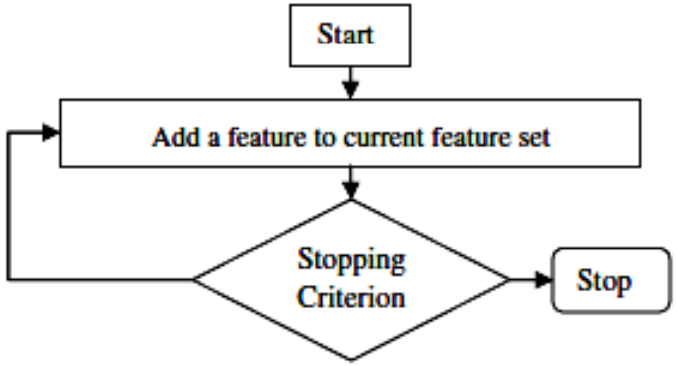
انتخاب ویژگی – ترکیب فیلتر و پوشش

- **بخش پوشش – ادامه**

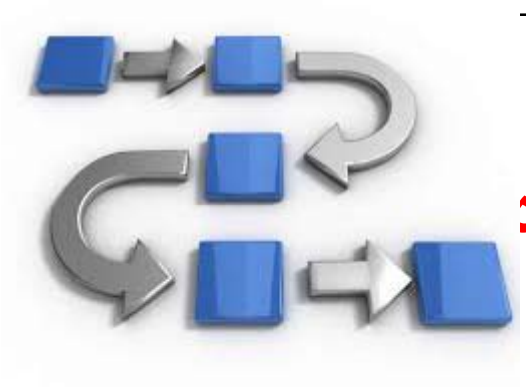
SBS



SFS



Hui-Huang Hsu, Cheng-Wei Hsieh, Ming-Da Lu, Hybrid feature selection by combining filters and wrappers, Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8144-8150



انتخاب ویژگی – روش تعبیه شده

- در این روش انتخاب ویژگی بعنوان مرحله ای از آموزش مدل یادگیری است و در آن تعبیه شده است

استخراج ویژگی

- در استخراج ویژگی می خواهیم k بعد جدید را با استفاده از d بعد اولیه ایجاد کنیم
- **دسته بندی روش ها**
- با ناظر یا بدون ناظر
- خطی یا غیرخطی
- ...

استخراج ویژگی

- **مفهوم تصویر[1]:** انتقال داده ها از یک فضا به فضایی با ابعاد کمتر به نحوی که کمترین میزان اطلاعات از بین رود

- تصویر $\boldsymbol{x}$ در جهت $\boldsymbol{w}$:

$$z = \boldsymbol{w}^T \boldsymbol{x}$$

1-projection

استخراج ویژگی

- **روش تحلیل اجزای اصلی[1]**

- بدون ناظر
- خطی

1-Principal component analysis

استخراج ویژگی

- **روش تحلیل اجزای اصلی**[1]

- داده ها به نحوی تصویر می شوند که **واریانس داده ها در فضای جدید حداکثر** شود.
- معادل با بردارهای ویژه ماتریس کوواریانس با مقادیر ویژه بیشتر

1-Principal component analysis

استخراج ویژگی

- **روش تحلیل اجزای اصلی[1]**

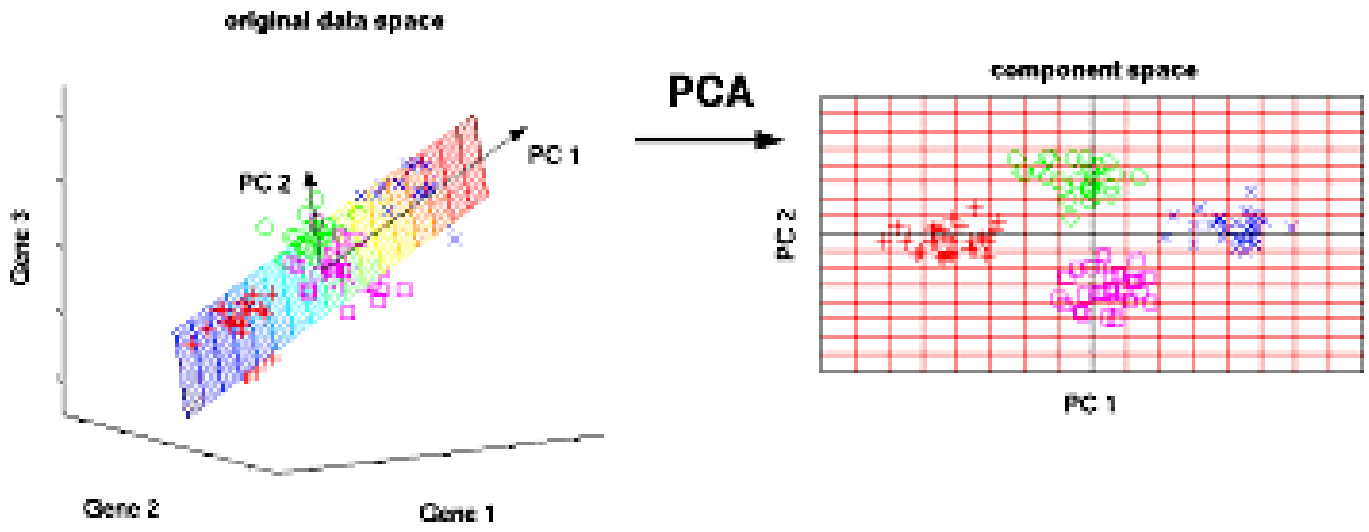
- اگر k بردار ویژه ماتریس کوواریانس با بیشترین مقادیر ویژه در ماتریس W باشند آنگاه در PCA نگاشت زیر انجام می شود

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

1-Principal component analysis

استخراج ویژگی

- روش تحلیل اجزای اصلی[1]



1-Principal component analysis

استخراج ویژگی

- **روش تحلیل جداسازی خطی[1]**

- با ناظر
- خطی

1-linear discriminant analysis

استخراج ویژگی

- **روش تحلیل جداسازی خطی[1]**

- اگر داده ها دو کلاسه باشند (C_1, C_2) ، نگاشتی را می خواهیم که پس از تصویر داده های دو کلاس تا حد امکان از هم جدا شوند.

1-linear discriminant analysis

استخراج ویژگی

- **روش تحلیل جداسازی خطی[1]**

- اگر میانگین داده ها در دو کلاس پس از تصویر m_1 و m_2 باشد و واریانس دو کلاس پس از تصویر s_1^2, s_2^2 در اینصورت در ***LDA*** نگاشتی را می یابد که عبارت زیر را بیشینه کند

$$J(w) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

1-linear discriminant analysis

استخراج ویژگی

- **روش تحلیل جداسازی خطی**[1]

$$J(\boldsymbol{w}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

- بیشینه شدن این عبارت **یعنی**

- ***بیشینه سازی واریانس بین کلاسی و کمینه سازی واریانس درون کلاسی***

1-linear discriminant analysis

استخراج ویژگی

- **روش تحلیل جداسازی خطی[1]**

- پس از مشتق‌گیری از $J(\boldsymbol{w}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$ داریم

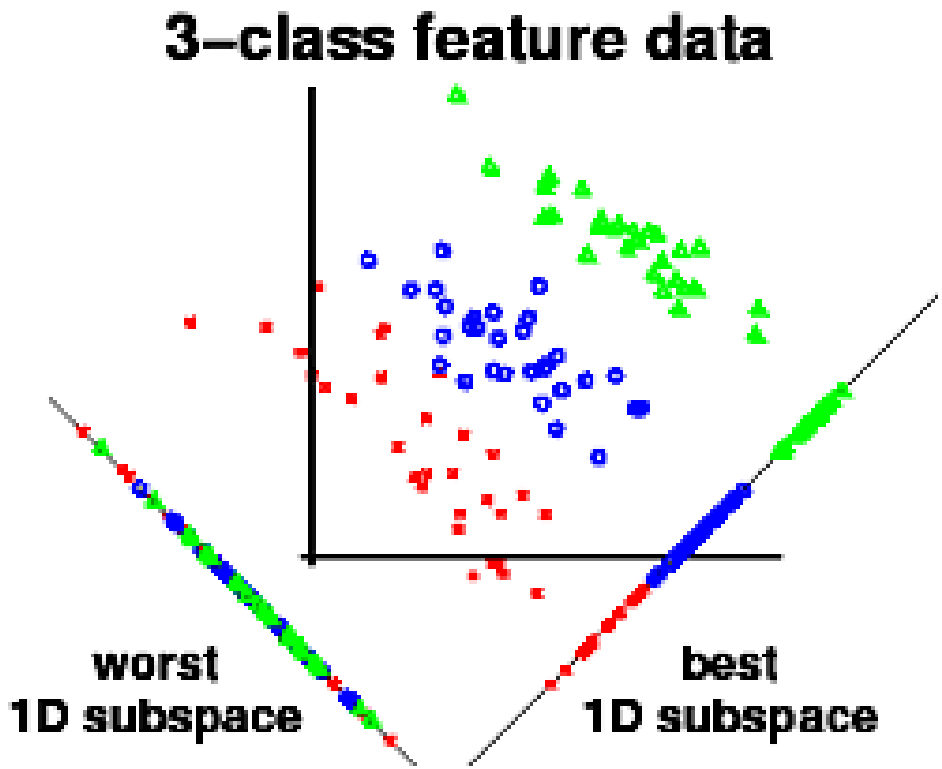
$$\boldsymbol{w} = c\mathbf{S}_W^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2)$$

- $\boldsymbol{Sw}$ واریانس درون کلاسی است

1-linear discriminant analysis

استخراج ویژگی

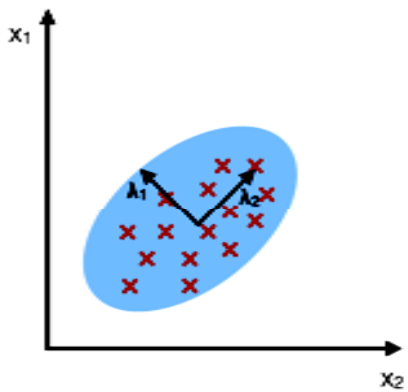
- روش تحلیل جداسازی خطی[1]



استخراج ویژگی

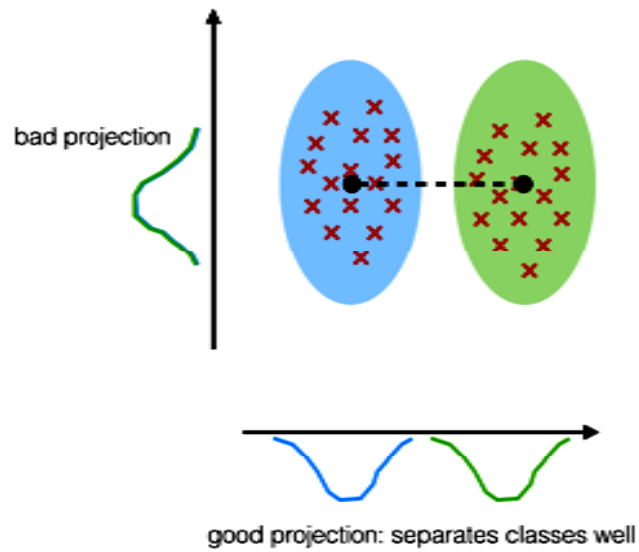
- مقایسه PCA و LDA

PCA:
component axes that maximize the variance



1

LDA:
maximizing the component axes for class-separation

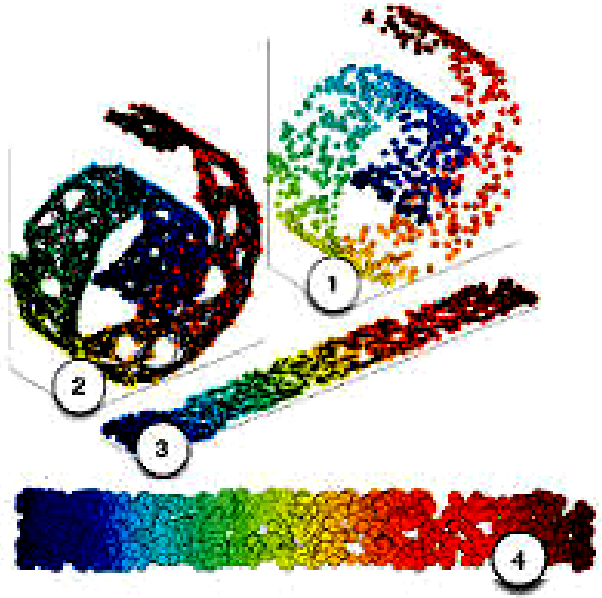


استخراج ویژگی

- **یادگیری مانیفولد**
- کاهش ابعاد غیرخطی

استخراج ویژگی

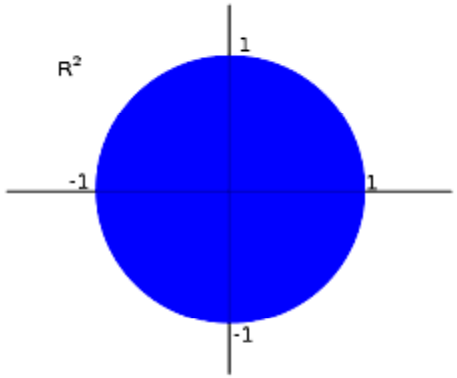
- **یادگیری مانیفولد**
- یادگیری نگاشت برای کاهش ابعاد



1-manifold learning

مانیفولد

- فضایی (مجموعه‌ای از نقاط) بهم مرتبط که بصورت محلی اقلیدسی است. یعنی اطراف هر نقطه همسایگی مشابه توپ واحد در $\boldsymbol{R}^n$ است



- **مثلاً:** در فضای دوبعدی $(\boldsymbol{R}^2)$

استخراج ویژگی

- **یادگیری مانیفولد**

- **روش‌ها**

  - LLE(Linear Local Embedding)
  - MVU (Maximum Variance Unfolding)
  - …

استخراج ویژگی

- **روش تعبیه محلی خطی[1]**

- در این روش تلاش می شود هر بخش از مانیفولد، بصورت خطی با استفاده از مجموع وزندار همسایگان تخمین زده شود

1-Linear Local Embedding

استخراج ویژگی

- **روش تعبیه محلی خطی[1]**

- اگر x^r , $x^s_{(r)}$ داده و همسایه های آن در فضای اصلی باشند، آنگاه وزنهای همسایه ها با مینیمم سازی رابطه زیر بدست می آید

$$\mathcal{E}^w(W|X) = \sum_r \|x^r - \sum_s W_{rs} x^s_{(r)}\|^2$$

1-Linear Local Embedding

استخراج ویژگی

- **روش تعبیه محلی خطی[1]**

- سپس از وزنهای W_{rs} برای پیدا کردن فضای کاهش یافته Z استفاده می کنیم

$$\mathcal{E}^{Z}(Z|\mathbf{W}) = \sum_{r} \|\mathbf{z}^{r} - \sum_{s} \mathbf{W}_{rs}\mathbf{z}^{s}\|^{2}$$

1-Linear Local Embedding

استخراج ویژگی

- **روش تعبیه محلی خطی[1]**
- بنابراین

$$X(\text{d-dimension}) \rightarrow Z(\text{k-dimension})$$

1-Linear Local Embedding

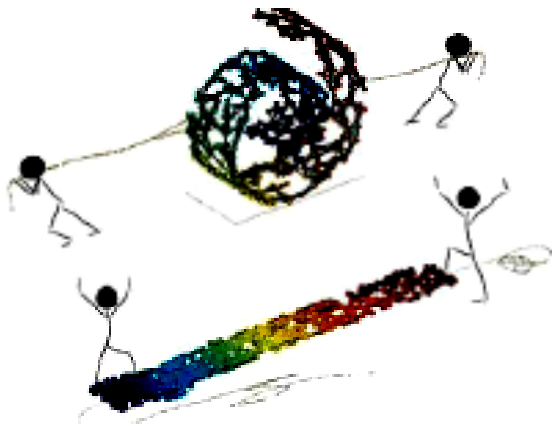
استخراج ویژگی

- **آشکار شدن مبتنی بر بیشترین واریانس[1]**

- فرض کنید هر نقطه در فضای اولیه بوسیله خطوطی به k همسایه نزدیک خود متصل باشد، در این الگوریتم سعی می‌شود مجموع فواصل نقاط حداکثر شود بدون اینکه خطوط ذکر شده شکسته شود

1-Maximum Variance Unfolding

استخراج ویژگی



- **آشکار شدن مبتنی بر بیشترین واریانس**
- **بیان ریاضی**

Maximize $\sum_{ij} \|\vec{y}_i - \vec{y}_j\|^2$ **subject to:**

(1) $\|\vec{y}_i - \vec{y}_j\|^2 = \|\vec{x}_i - \vec{x}_j\|^2$ **for all** (i, j) **with** $\eta_{ij} = 1$.

(2) $\sum_i \vec{y}_i = 0$

- y فضای جدید، x فضای اولیه $\eta_{ij} \in \{0, 1\}$ آیا نقاط در فضای اولیه جزو k نزدیکترین همسایه هستند

جمع بندی

- کاهش ابعاد در دو شکل انتخاب ویژگی و استخراج ویژگی
- در هر دسته روشهای متنوع وجود دارد