

فرادرس

فراتر از یک کلاس درس
www.faradars.org

آموزش یادگیری ماشین (Machine Learning)

(تئوری - عملی) - بخش دوم

درس چهارم: انتخاب ویژگی

مدرس:

فرشید شیرافکن

دانشجوی دکترای بیو انفورماتیک

دانشگاه تهران

انتخاب ویژگی

انتخاب یک زیرمجموعه از مجموعه ویژگی های داده شده.

	1	2	3	4	-	-	d
1							
2							
\vdots	\vdots	\vdots			\dots		\vdots
100							
	i_1			i_2	\dots		$i_{d'}$
	\uparrow			\uparrow			\uparrow

$$d = 100$$

$$d' = 20$$

$$d' < d$$

$$[x_1, x_2, \dots, x_d]$$



$$[x_{i_1}, x_{i_2}, \dots, x_{i_{d'}}]$$

انتخاب ویژگی

- انتخاب زیر مجموعه ای از ویژگی ها لازم دارد:

۱- یک استراتژی جستجو برای انتخاب زیر مجموعه های کاندید.

۲- یک تابع هدف (objective function) برای ارزیابی این کاندیدها.

تک متغیره (Univariate method):

در یک زمان فقط یک ویژگی بررسی می شود. (هر ویژگی را مستقل از سایر ویژگی ها در نظر می گیرد)

چند متغیره (Multivariate method):

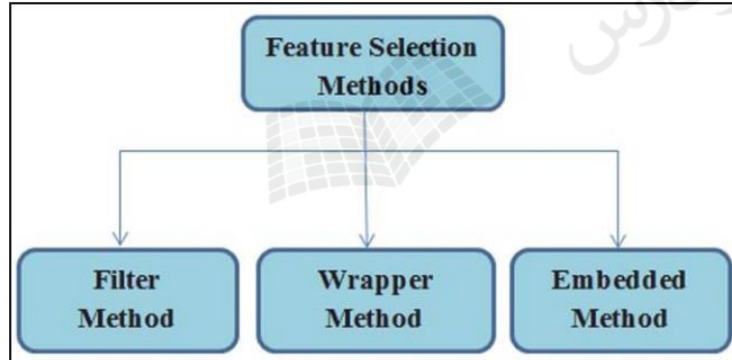
زیر مجموعه ای از ویژگی ها با هم بررسی می شوند.

روش‌های انتخاب ویژگی

۱- روش فیلتر (Filter method)

۲- روش بسته‌بند (Wrapper method)

۳- روش توکار (Embedded method)



روش فیلتر (Filter method)

✓ ویژگی‌ها مستقل از دسته‌بند (به عنوان مرحله پیش پردازش) رتبه‌بندی می‌شوند.

✓ توسط یک اندازه‌گیری، امتیازی به ترکیب‌های مختلف ویژگی‌ها بدون در نظر گرفتن الگوریتم دسته‌بندی، نسبت داده می‌شود.



- ✓ اغلب تک متغیره.
- ✓ ساده و سریع.
- ✓ دیتاست‌های با ابعاد بالا از روش فیلتر استفاده می‌کنند.
- ✓ معیارهای اندازه‌گیری: همبستگی - اطلاعات متقابل.

Correlation

$$R(k) = \frac{\text{cov}(X_k, Y)}{\sqrt{\text{var}(X_k)}\sqrt{\text{var}(Y)}}$$

✓ مقدار correlation بین هر ویژگی با خروجی را محاسبه می کنیم.
✓ ویژگی با correlation بیشتر، مناسب تر است.

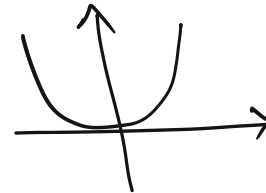
$$X = (x_1, x_2, x_3, x_4)$$

$$\begin{cases} \text{cor}(x_1, y) \\ \text{cor}(x_2, y) \\ \text{cor}(x_3, y) \\ \text{cor}(x_4, y) \end{cases}$$

* ایراد

• وابستگی های غیر خطی را نمی تواند تشخیص دهد.

$$y = x^2$$



Mutual Information

MI: به عنوان یک معیار وابستگی

$$MI(X_1, Y)$$

$$MI(X_2, Y)$$

⋮

مقدار MI بیشتر: وابستگی بیشتر بین ویژگی با برچسب.

✓ ویژگی های وابسته به برچسب، مناسب ترند.

هر اندازه که مستقل نباشند، یعنی X درباره Y اطلاع می دهد.

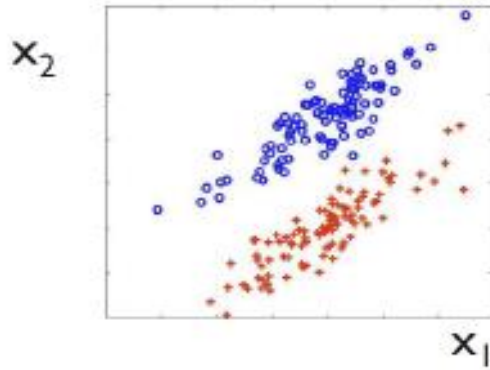
$$MI(X, Y) = E_{X, Y} \left[\log \frac{P(X, Y)}{P(X)P(Y)} \right]$$

$$P(X, Y) = P(X)P(Y) \quad \text{مستقل}$$

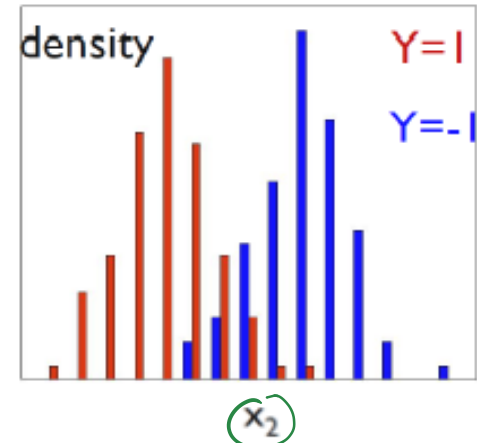
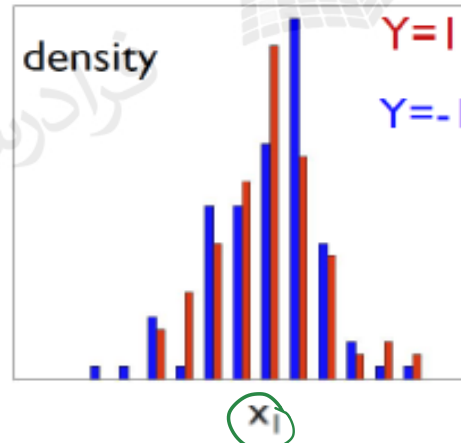
$$MI = 0$$

$$I(k) = MI(X_k, Y)$$

مثال



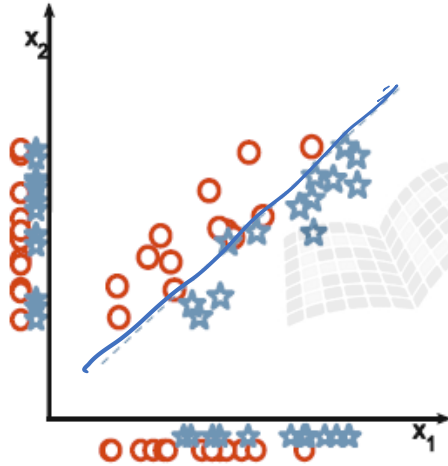
$(x_1, x_2) \rightarrow ? \quad x_2$



$$P(X_1 | Y = 1) = P(X_1 | Y = -1)$$

معایب تک متغیره

- شاید یک ویژگی به تنهایی مناسب نباشد، ولی ترکیب آن با ویژگی دیگر باعث دسته‌بندی با دقت خوبی شود.
- در شکل زیر، هر کدام از ویژگی‌ها به تنهایی نمی‌توانند تفکیک خوبی بین دو دسته ایجاد کنند، اما ترکیب هر دو خوب است و دو دسته با یک خط به خوبی جدا می‌شود.



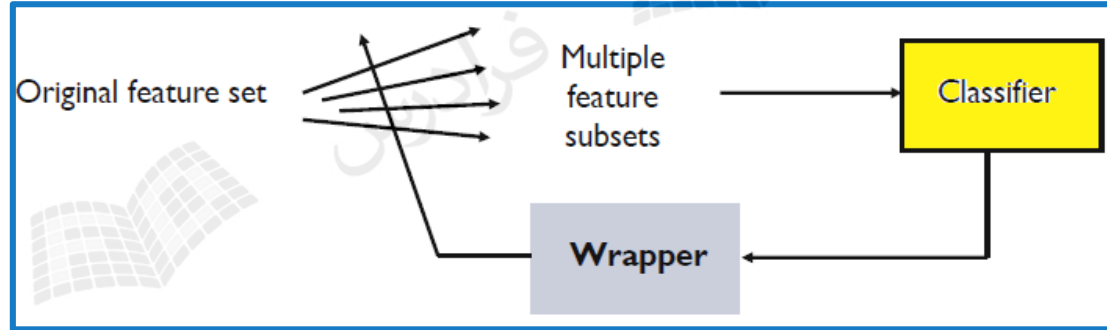
x_1

x_2

(x_1, x_2)

روش بسته‌بند (Wrapper method)

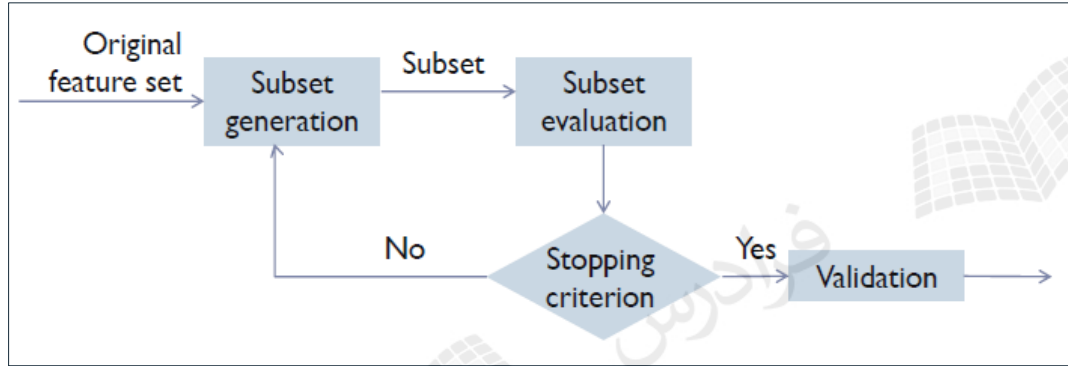
- استفاده از دسته‌بند برای ارزیابی ویژگی‌ها.
- برای هر زیرمجموعه از ویژگی‌ها، دسته‌بند بر روی داده‌های آموزشی، train شده و کارایی آن با استفاده از تکنیک‌های ارزیابی شبیه cross-validation ارزیابی می‌شود.



- Filter methods use statistical methods for evaluation of a subset of features while wrapper methods use cross validation.

رویه عمومی

تولید زیر مجموعه و ارزیابی آن تا زمانی که شرط توقف رخ دهد.



معیارهای توقف:

- پیش تعریف تعداد ویژگی‌های انتخابی یا تعداد تکرارها
- به دست آوردن یک زیر مجموعه بهینه (طبق معیارهای ارزیابی)
- اضافه کردن یک ویژگی باعث بهبود نشود.

مقایسه

Filters

Methods:

- Criterion: Measure feature/feature subset "relevance"
- Search: Usually order features (individual feature ranking or nested subsets of features)
- Assessment: Use statistical tests

Results:

- Are (relatively) robust against overfitting
- May fail to select the most "useful" features

Generality
بی‌محدودیت
سرریز

Wrappers

Methods:

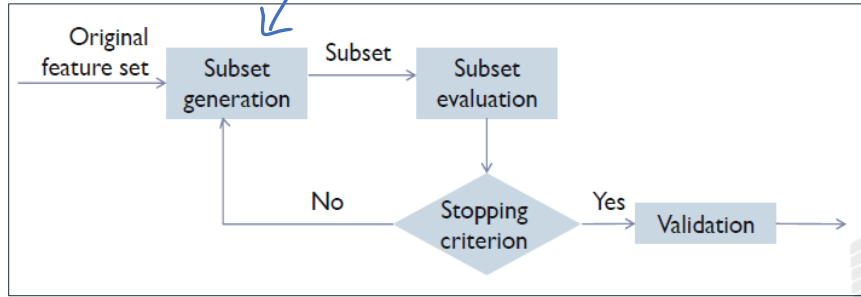
- Criterion: Measure feature subset "usefulness"
- Search: Search the space of all feature subsets
- Assessment: Use cross-validation

Results:

- Can in principle find the most "useful" features, but
- Are prone to overfitting

دقت

استراتژی‌های جستجو



۱- کامل (Exhaustive)

$O(2^n)$
B & B ← حیار فاصله

۲- تصادفی (Probabilistic)

L_{SW} ← نرخ خطا طبقه‌بندی
 L_{SF} ← consistency

۳- مکاشفه ای (Heuristic)

SFS ← نرخ خطا کلاسی فایر
 SBS

information ← DTM

evaluation measures :

- distance
- information
- dependence
- Consistency
- classifier error rate

$\{f_1, f_2, f_3\}$

$\{\{f_1\}, \{f_2\}, \{f_3\}\}$

$\{f_1, f_2\}, \{f_1, f_3\}, \{f_2, f_3\}$

$\{f_1, f_2, f_3\}$

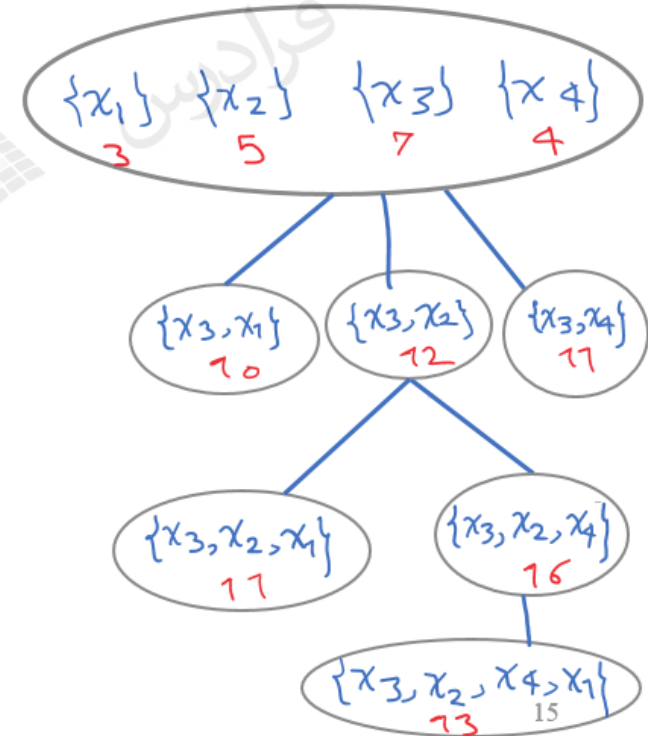
استراتژی‌های جستجو

- Sequential forward selection (SFS)
- Sequential backward selection (SBS)
- Plus-L minus-R selection (LRS)
- Bidirectional search
- floating search
- ...

Sequential Forward Selection (SFS)

1. Start with the empty set $Y_0 = \{\emptyset\}$
2. Select the next best feature $x^+ = \arg \max_{x \notin Y_k} J(Y_k + x)$
3. Update $Y_{k+1} = Y_k + x^+; k = k + 1$
4. Go to 2

k	Y_k
0	$\{\emptyset\}$
1	$\{x_3\}$
2	$\{x_3, x_2\}$
3	$\{x_3, x_2, x_4\}$



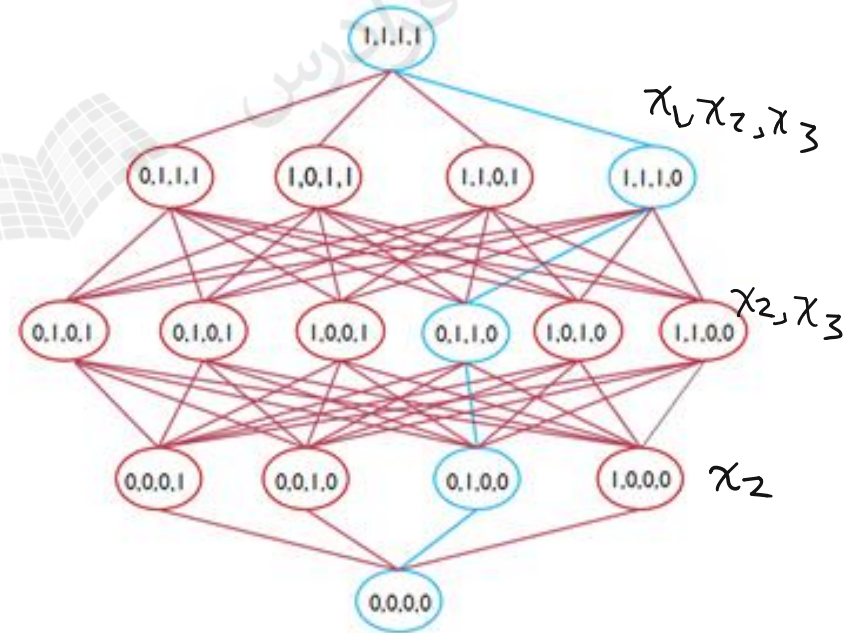
GSFS →

Sequential Backward Selection (SBS)

GSBS →

1. Start with the full set $Y_0 = X$
2. Remove the worst feature $x^- = \arg \max_{x \in Y_k} J(Y_k - x)$
3. Update $Y_{k+1} = Y_k - x^-; k = k + 1$
4. Go to 2

SFS



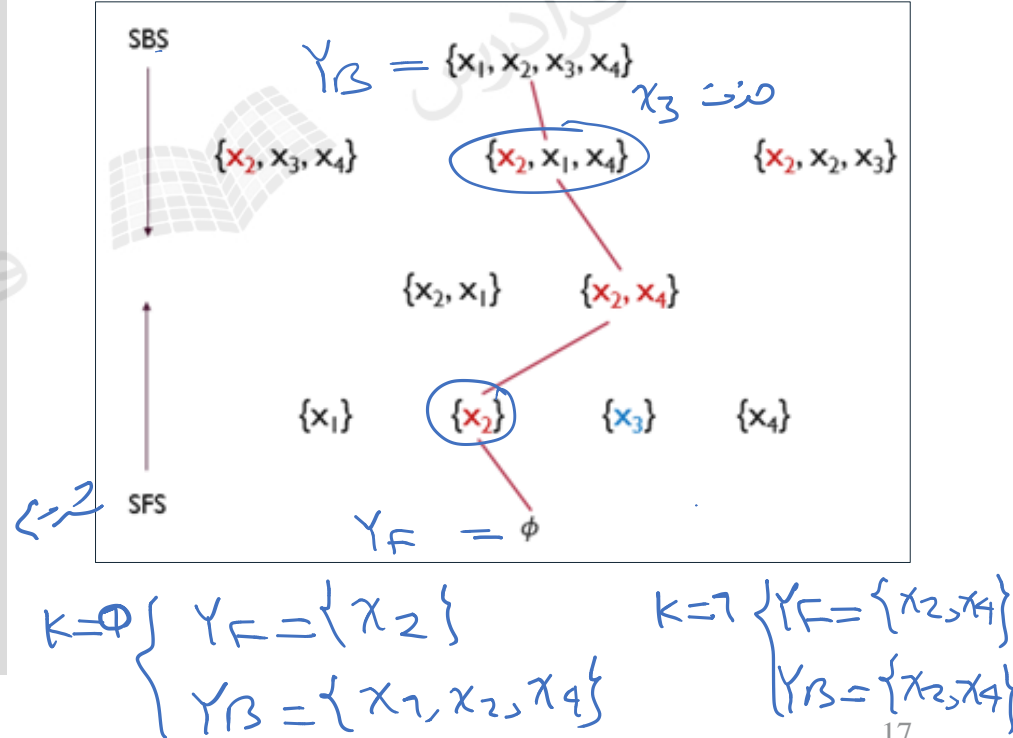
Bidirectional Search (BDS)

- 1. Start SFS with $Y_F = \{\emptyset\}$
- 2. Start SBS with $Y_B = X$
- 3. Select the best feature

$$\begin{cases} x^+ = \arg \max_{\substack{x \notin Y_{F_k} \\ x \in F_{B_k}}} J(Y_{F_k} + x) \\ Y_{F_{k+1}} = Y_{F_k} + x^+ \end{cases}$$

x_2
4. Remove the worst feature

$$\begin{cases} x^- = \arg \max_{\substack{x \in Y_{B_k} \\ x \notin Y_{F_{k+1}}}} J(Y_{B_k} - x) \\ Y_{B_{k+1}} = Y_{B_k} - x^-; \underline{k = k + 1} \end{cases}$$
5. Go to 3



Plus-L Minus-R Selection (LRS)

1. If $\underline{L} > R$ then $Y_0 = \{\emptyset\}$
 else $Y_0 = X$; go to step 3
2. Repeat L times

$$\begin{cases} x^+ = \arg \max_{x \notin Y_k} J(Y_k + x) \\ Y_{k+1} = Y_k + x^+; k = k + 1 \end{cases}$$
3. Repeat R times

$$\begin{cases} x^- = \arg \max_{x \in Y_k} J(Y_k - x) \\ Y_{k+1} = Y_k - x^-; k = k + 1 \end{cases}$$
4. Go to 2

$L = 4$ $R = 2$ $L > R$ $Y_0 = \{\emptyset\}$ L مرتبه اضافه R مرتبه حذف	$L = 3$ $R = 7$ $L < R$ $Y_0 = X$ R مرتبه حذف L مرتبه اضافه
--	--

Sequential Floating Forward Selection (SFFS)

1. $Y = \{\emptyset\}$
2. Select the best feature
$$x^+ = \arg \max_{x \notin Y_k} J(Y_k + x)$$
$$Y_k = Y_k + x^+; k = k + 1$$
3. Select the worst feature
$$x^- = \arg \max_{x \in Y_k} J(Y_k - x)$$
4. If $J(Y_k - x^-) > J(Y_k)$ then
$$Y_{k+1} = Y_k - x^-; k = k + 1$$

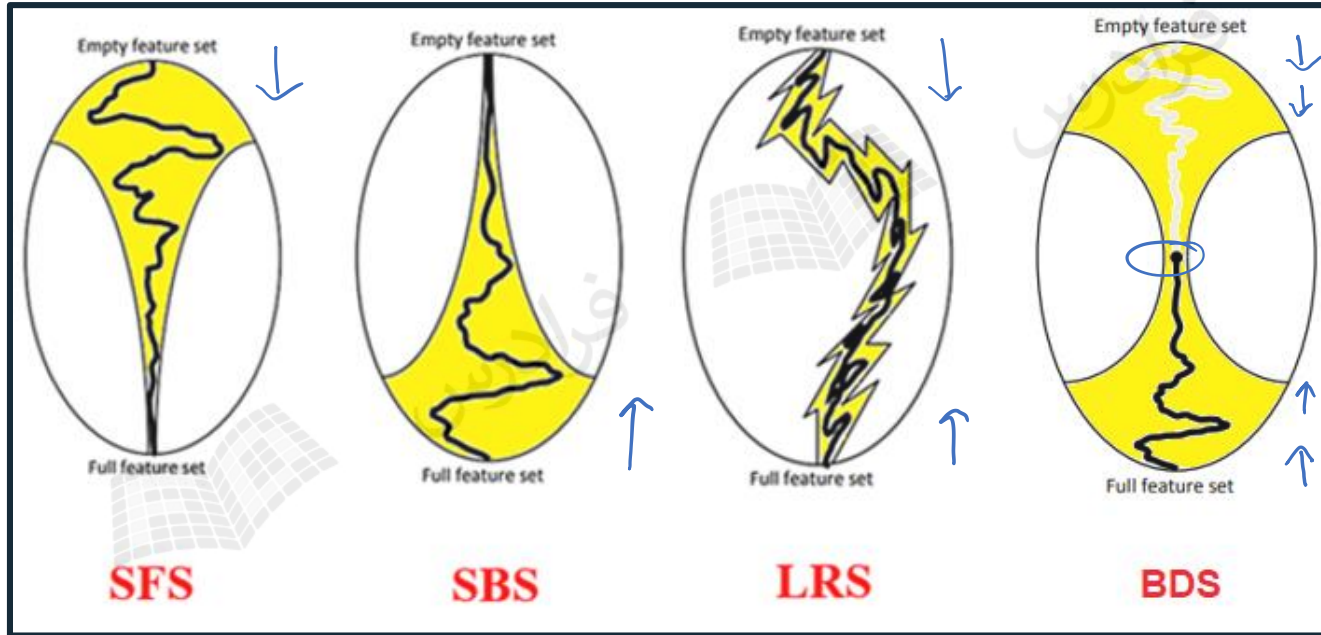
Go to step 3

Else

Go to step 2

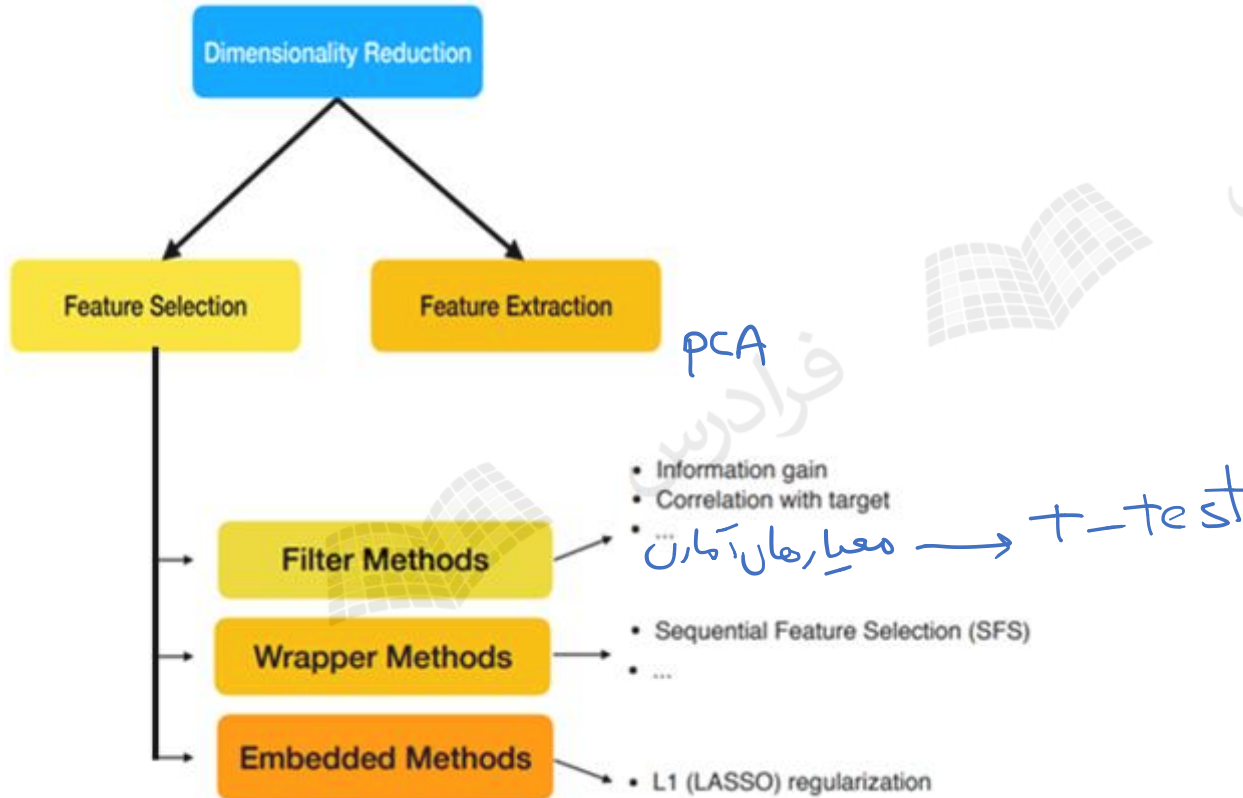
LRS

$L > R$



$L > R$
 $L < R$

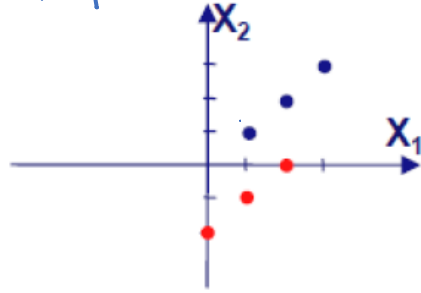
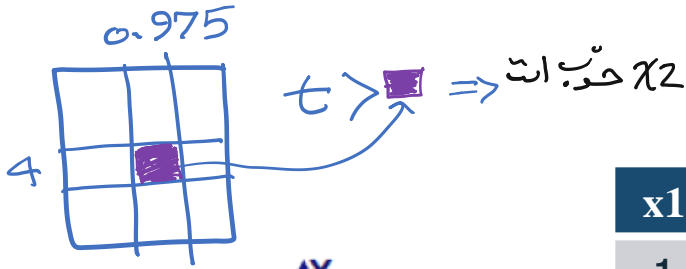
جمع بندی



آزمون آماری t

$$\alpha = \text{سطح اطمینان} = 0.05$$

$$1 - \frac{\alpha}{2} = 0.975$$



x1	x2
1	1
2	2
3	3
0	-2
1	-1
2	0

آبی

قرمز

$$t = \frac{2 - (-1)}{\sqrt{\frac{1}{3} + \frac{1}{3}}} = 3.6$$

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

$$df = \frac{(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y})^2}{\frac{(\frac{s_x^2}{n_x})^2}{n_x - 1} + \frac{(\frac{s_y^2}{n_y})^2}{n_y - 1}}$$

ویرتی x_2

میانگین

آبی

قرمز

2

-1

واریانس

آبی

قرمز

6

1

$$df = \frac{(\frac{1}{3} + \frac{1}{3})^2}{(\frac{1}{3})^2/2 + (\frac{1}{3})^2/2} = \frac{4}{9} = 4$$

درج آزادی

انتخاب ویژگی با الگوریتم های فرا ابتکاری

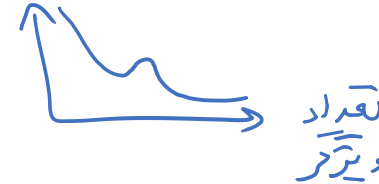
انتخاب ویژگی یک مسئله بهینه سازی است

یک هدف:

$$Z = E(1 + Kr)$$

درصد ویژگی های صریح مقدار انتخاب شده

E



$$0 \leq r \leq 1$$

Meta_heuristic

Evolutionary

GA

Swarm intelligence

PSO

بهینه سازی ازدحام ذرات

Physics_based

SA

شبیه سازی تبرید

$$e = |t - y|$$

x ویژگی

FS

\hat{x}

تاریک t → فزاینده

خروجی y → مدل

این اسلایدها بر مبنای نکات مطرح شده در فرادرس
«آموزش یادگیری ماشین (Machine Learning) (تئوری - عملی) - بخش دوم»
تهیه شده است.

برای کسب اطلاعات بیشتر در مورد این آموزش به لینک زیر مراجعه نمایید.

faradars.org/fvdm94062