

회귀분석

엑셀 추세선 유사

Why : 현재 선형시각화 + 미래추세 가늠

1. 선형회귀 분석

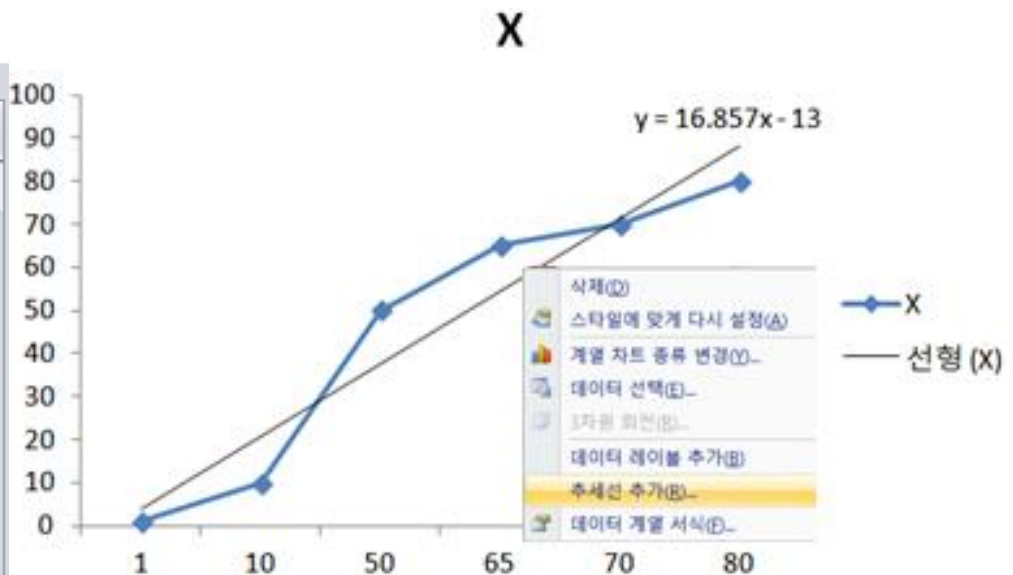
- 일반적인 $y = Ax+B$ 이며, x 와 y 는 숫자형
- x 의 개수는 많아도 동일형태 ($y = A_1x_1+A_2x_2+B$)

숫자형

Viewer

Relation: cpu

No.	1: MYCT	2: MMIN	3: MMAX	4: CACH	5: CHMIN	6: CHMAX	7: class
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
181	160.0	2000.0	8000.0	32.0	1.0	13.0	109.0
182	240.0	512.0	1000.0	8.0	1.0	3.0	6.0
183	240.0	512.0	2000.0	8.0	1.0	5.0	11.0
184	105.0	2000.0	4000.0	8.0	3.0	8.0	22.0
185	105.0	2000.0	6000.0	16.0	6.0	16.0	33.0
186	105.0	2000.0	8000.0	16.0	4.0	14.0	58.0
187	52.0	4000.0	1600...	32.0	4.0	12.0	130.0
188	70.0	4000.0	1200...	8.0	6.0	8.0	75.0
189	59.0	4000.0	1200...	32.0	6.0	12.0	113.0
190	59.0	8000.0	1600...	64.0	12.0	24.0	188.0
191	26.0	8000.0	2400...	32.0	8.0	16.0	173.0
192	26.0	8000.0	3200...	64.0	12.0	16.0	248.0
193	26.0	8000.0	3200...	128.0	24.0	32.0	405.0



2. M5P

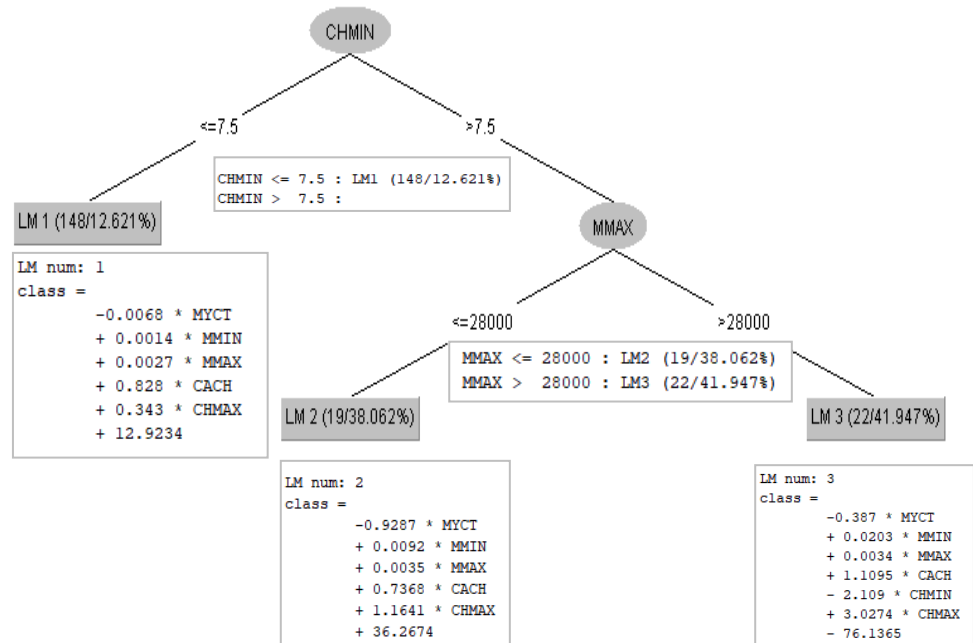
- 의사결정나무와 회귀분석 동시 학습
- 분류규칙은 노드간 분기조건, 회귀식은 말단노드 조건

숫자형

Viewer

Relation: cpu

No.	1: MYCT	2: MMIN	3: MMAX	4: CACH	5: CHMIN	6: CHMAX	7: class
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
181	160.0	2000.0	8000.0	32.0	1.0	13.0	109.0
182	240.0	512.0	1000.0	8.0	1.0	3.0	6.0
183	240.0	512.0	2000.0	8.0	1.0	5.0	11.0
184	105.0	2000.0	4000.0	8.0	3.0	8.0	22.0
185	105.0	2000.0	6000.0	16.0	6.0	16.0	33.0
186	105.0	2000.0	8000.0	16.0	4.0	14.0	58.0
187	52.0	4000.0	1600...	32.0	4.0	12.0	130.0
188	70.0	4000.0	1200...	8.0	6.0	8.0	75.0
189	59.0	4000.0	1200...	32.0	6.0	12.0	113.0
190	59.0	8000.0	1600...	64.0	12.0	24.0	188.0
191	26.0	8000.0	2400...	32.0	8.0	16.0	173.0
192	26.0	8000.0	3200...	64.0	12.0	16.0	248.0
193	26.0	8000.0	3200...	128.0	24.0	32.0	405.0



3. 선형회귀 분석 (2라벨)

- $y = Ax+B$ 中 y 의 데이터 값이 2가지로 분류 (yes/no)
- 명목형 y 의 데이터 값들을 이진숫자형으로 변환 (0/1)
(필터 NorminalToBinary)

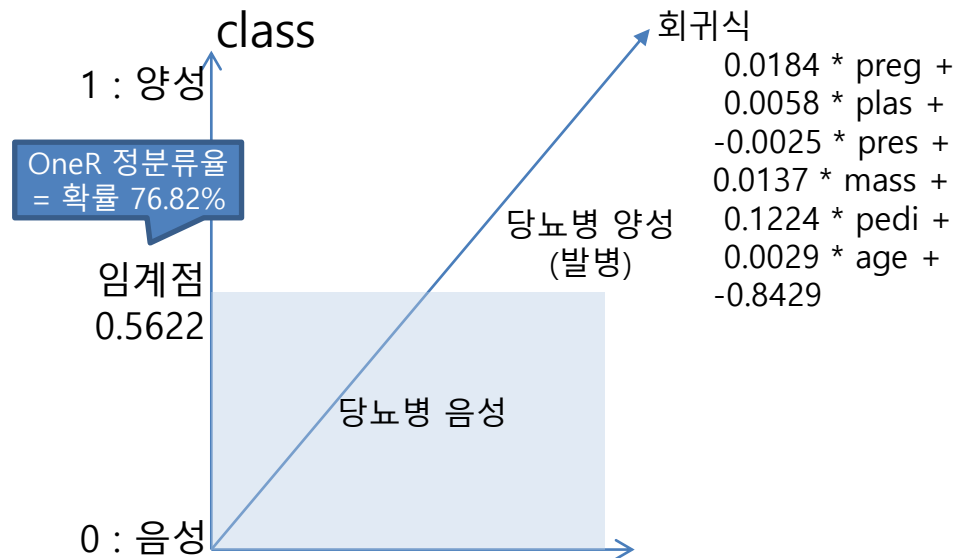
Relation: pima_diabetes

No.	1: preg	2: plas	3: pres	4: skin	5: insu	6: mass	7: pedi	8: age	9: class
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
...	2.0	92.0	52.0	0.0	0.0	30.1	0.141	22.0	tested_negative
...	2.0	106.0	56.0	27.0	165.0	29.0	0.426	22.0	tested_negative
...	2.0	105.0	75.0	0.0	0.0	23.3	0.56	53.0	tested_negative
...	4.0	95.0	60.0	32.0	0.0	35.4	0.284	28.0	tested_negative
...	0.0	126.0	86.0	27.0	120.0	27.4	0.515		
...	8.0	65.0	72.0	23.0	0.0	32.0	0.6		

Viewer

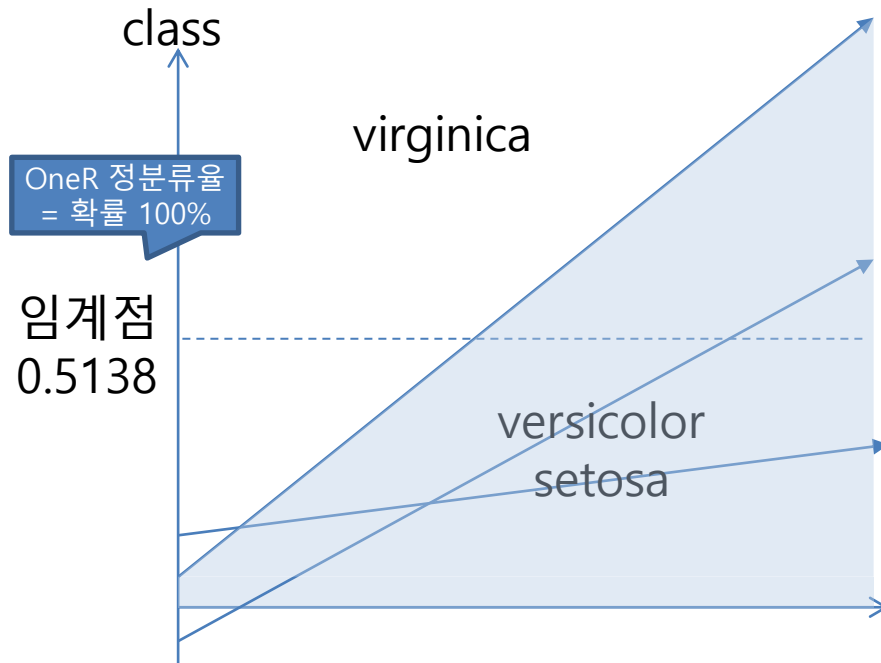
Relation: pima_diabetes-weka.filters.unsupervised.attribute.NominalToBinary-k1ast

No.	1: preg	2: plas	3: pres	4: skin	5: insu	6: mass	7: pedi	8: age	9: class=tested_positiv
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	6.0	148.0	72.0	35.0	0.0	33.6	0.627	50.0	1.0
2	1.0	85.0	66.0	29.0	0.0	26.6	0.351	31.0	0.0
3	8.0	183.0	64.0	0.0	0.0	23.3	0.672	32.0	1.0
4	1.0	89.0	66.0	23.0	94.0	28.1	0.167	21.0	0.0
5	0.0	137.0	40.0	35.0	168.0	43.1	2.288	33.0	1.0
6	5.0	116.0	74.0	0.0	0.0	25.6	0.201	30.0	0.0
7	3.0	78.0	50.0	32.0	88.0	31.0	0.248	26.0	1.0
8	10.0	115.0	0.0	0.0	0.0	35.3	0.134	29.0	0.0
9	2.0	197.0	70.0	45.0	543.0	30.5	0.158	53.0	1.0
10	8.0	125.0	96.0	0.0	0.0	0.0	0.232	54.0	1.0



4. 선형회귀분석 (3라벨 이상)

- $y = Ax+B$ 中 y 데이터 값이 3가지로 이상 분류 (상/중/하, 1~10 등급)
- 명목형 y 의 데이터 값들을 숫자형으로 변환 (필터 : MakeIndicator)
- 라벨개수만큼 생성된 회귀식별 **최고 상관계수 식**을 채택



Indice 1 (setosa) 상관계수 **0.9456 (최대출력 채택)**

$$\begin{aligned} \text{class} = & 0.0741 * \text{sepalwidth} + \\ & 0.2445 * \text{sepalwidth} + \\ & -0.2117 * \text{petallength} + \\ & -0.0941 * \text{petalwidth} + \\ & 0.0591 \end{aligned}$$

Indice last (virginica) 상관계수 0.7676

$$\begin{aligned} \text{class} = & 0.175 * \text{sepalwidth} + \\ & 0.518 * \text{petalwidth} + \\ & -0.8162 \end{aligned}$$

Indice 2 (veisicolor) 상관계수 0.458

$$\begin{aligned} \text{class} = & -0.4633 * \text{sepalwidth} + \\ & 0.1856 * \text{petallength} + \\ & -0.4421 * \text{petalwidth} + \\ & 1.5691 \end{aligned}$$

5. 로지스틱 회귀분석 (도 아니면 모)

- 로짓변환으로 $y = Ax + B$ 를 $y = \exp(Cx + d)$ 로 변형
- 지수함수(exp)를 사용하므로 명목목적변수 분류/예측 用

명목형
tested_positive
tested_negative

$$\text{로지스틱회귀식} = e^{(-0.123 * \text{preg} - 0.035 * \text{plas} + 0.013 * \text{pres} + 8.4047)}$$

Relation: pima_diabetes									
No.	1: preg	2: plas	3: pres	4: skin	5: insu	6: mass	7: pedi	8: age	9: class
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
...	2.0	92.0	52.0	0.0	0.0	30.1	0.141	22.0	tested_negative
...	2.0	106.0	56.0	27.0	165.0	29.0	0.426	22.0	tested_negative
...	2.0	105.0	75.0	0.0	0.0	23.3	0.56	53.0	tested_negative
...	4.0	95.0	60.0	32.0	0.0	35.4	0.284	28.0	tested_negative
...	0.0	126.0	86.0	27.0	120.0	27.4	0.515	21.0	tested_negative
...	8.0	65.0	72.0	23.0	0.0	32.0	0.6	42.0	tested_negative
...	2.0	99.0	60.0	17.0	160.0	36.6	0.453	21.0	tested_negative
...	3.0	102.0	44.0	20.0	94.0	30.8	0.4	26.0	tested_negative
...	1.0	109.0	58.0	18.0	116.0	28.5	0.219	22.0	tested_negative
...	13.0	153.0	88.0	37.0	140.0	40.6	1.174	39.0	tested_negative
...	12.0	100.0	84.0	33.0	105.0	30.0	0.488	46.0	tested_negative
...	1.0	81.0	74.0	41.0	57.0	46.3	1.096	32.0	tested_negative
...	1.0	121.0	78.0	39.0	74.0	39.0	0.261	28.0	tested_negative

