

[다변량데이터분석] 과제 #3

CART & Neural Network

2018170809 오민제

Q1

데이터셋 #1 선정

의사와 데이터과학자들의 협력을 통해 헬스케어 분야에서 데이터과학을 접목시키려는 시도는 최근에 급증하고 있다. 이때 EMR(전자건강기록)을 바탕으로 머신 러닝 모델을 구축하게 되는데, 헬스케어 분야 특성상 왜 그런 결과를 도출하게 되었는지를 비전공자인 의사에게 설명해야 하며, 헬스케어가 생명과 직결된 분야인만큼 신뢰도가 중요한데, 이를 위해서는 모델의 해석 여부가 중요하다. 본 보고서에서는 여러 EMR 중 심근경색 데이터를 활용해 여러 변수들을 통해 심장마비 여부를 예측하는 모델을 구축하는 것을 목표로 한다. 본 데이터셋의 크기가 303으로 작은 편에 속하기 때문에 4:2:4 (학습:검증:테스트)로 데이터셋을 나누어 테스트 데이터를 충분히 확보하고 모델을 구축하겠다.

Feature	Description	Feature	Description
age	Age in years	thalach	
sex	Gender (0=M, 1=F)	exang	Exercise induced angina
cp	Chest pain type	oldpeak	ST depression induced by exercise relative to rest
trestbps	Resting blood pressure	slope	The slope of the peak exercise ST segment
chol	Serum cholestrol	ca	Number of major vessels
fbs	Fasting blood sugar	thal	Absorbability of thallium
restecg	Resting electrocardiogram results	target	Heart Attack (target feature)

출처 <https://www.kaggle.com/datasets/pritsheta/heart-attack>

Table 1. 데이터셋 #1의 주요 변수 설명 (N=303)

Q2

CART

우선 모든 학습+검정 데이터를 활용해 Full Tree를 구축하고, 이후 검증 데이터를 활용해 Post Pruning을 진행하도록 하겠다. 이때 (학습데이터+검증데이터):(검증 데이터)는 60:20이므로 K=3인 k-fold 방식을 통해서 오분류율이 최소가 되도록 Pruning을 진행한다. Fig 1.을 통해서 Pruning된 최적의 트리는 리프노드가 5개일때라는 것을 알 수 있다.

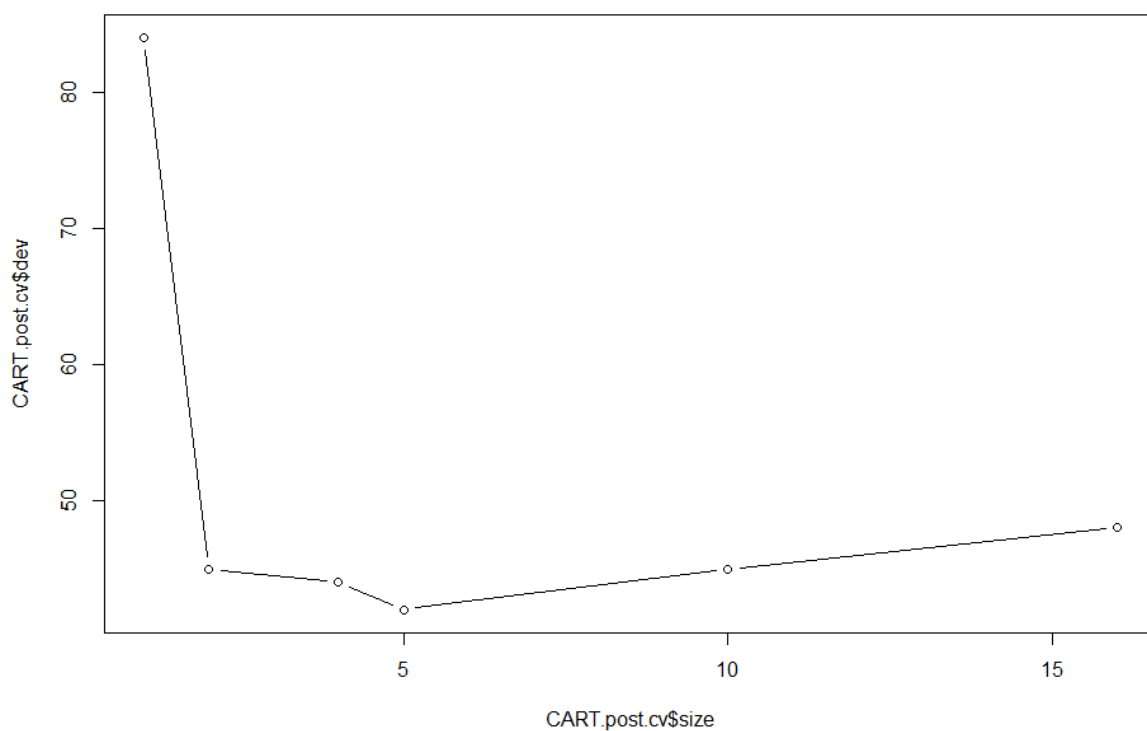


Fig 1. 검증 데이터에 대한 리프 노드별 오분류율

	TPR	Precision	TNR	Accuracy	BCR	F1-measure	Rank
Full Tree	0.8060	0.7826	0.7222	0.7686(2)	0.7630(1)	0.7941(2)	1.67
Post-Pruned	0.8955	0.7500	0.6296	0.7769(1)	0.7509(2)	0.8163(1)	1.33

Table 2. Full Tree 와 Post-Pruning Tree의 주요 이진분류 성능 지표들 비교

두 트리를 테스트 데이터셋을 통해 비교한 결과, Post-Pruning된 트리가 전반적으로 우월한 성능을 보임을 확인할 수 있다.

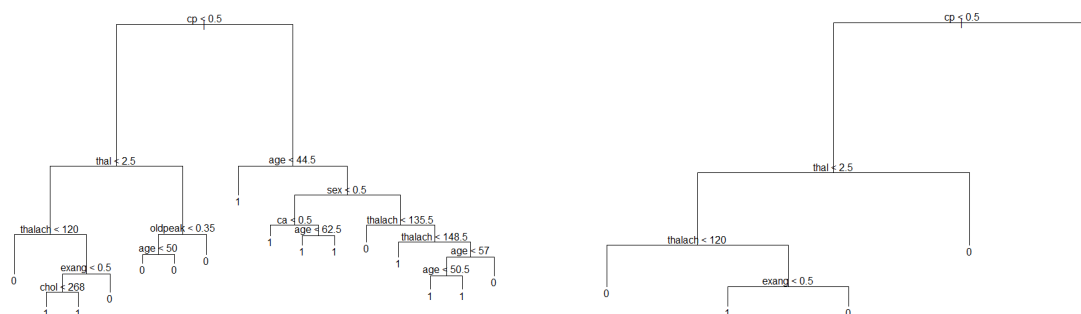


Fig 2. Full Tree(좌) 와 Post-Pruned Tree(우) 비교

Confusion Matrix		Predicted	
		0	1
Actual	0	39	15
	1	13	54

Confusion Matrix		Predicted	
		0	1
Actual	0	34	20
	1	7	60

Table 3. 테스트 데이터에 대한 Full Tree(좌)와 Post-Pruned Tree(우)의 confusion matrix

Q3

Pre-Pruning

학습데이터와 검증 데이터를 활용해 Pre-Pruning을 진행한다. 이 과정에서 사용되는 하이퍼파라미터는 다음과 같다. **Min.criterion**은 split의 불순도 감소의 통계적 유의성을 확인해 최소 신뢰도를 정하는 것으로, 통상 사용되는 신뢰도 기준인 {0.900, 0.950, 0.990, 0.995}를 사용한다. **Min.split**은 한 노드가 split되기 위해 필요한 최소 관측치의 개수를 의미하며, 본 데이터셋이 그리 크지 않기 때문에, {5,10,30}을 사용한다. **Max.depth**는 트리의 깊이에 대한 상한을 지정하는 것으로 Q2의 Full Tree가 depth = 5였음을 기억한다면 {0,3,4,5}를 사용한다. 이때 가능한 총 48개의 하이퍼파라미터 조합을 AUROC로 비교 시 **min.criteria** = 0.9, **min.split** = 5, **max.depth** = 0 일 때 AUROC = 0.97333로 최적임을 알 수 있다. 최적의 하이퍼파라미터를 바탕으로 Pre-Pruning된 Decision Tree를 구축한다. (*하이퍼파라미터 조합은 supplementary 참고)

Confusion Matrix		Predicted	
		0	1
Actual	0	33	21
	1	8	59

Table 4. 테스트 데이터에 대한 Pre-Pruned Tree의 Confusion Matrix

Q4 & Q5

Pruning 비교

Fig 2. 에서 Post-Pruned 트리의 경우 **cp**가 0.5 초과라면 심근경색으로 판별, 이하라면 **thal**을 확인해 2.5 초과라면 건강한 것으로 판별한다. 만약 **thal**이 2.5이하라면 **thalach**을 확인하고 120 이하라면 건강한 것으로 판별하고, 120 초과라면 **exang**을 확인해 0.5 이하라면 심근경색으로 판별, 0.5 초과라면 건강한 것으로 판별한다.

Fig 3. 에서 Pre-Pruned 트리의 경우 **cp**가 0 초과라면 **sex**를 확인하고 0 이하라면(남자) 심근경색으로 판별하고, 0 초과라면(여자) **thalach**을 확인해 134 초과이면 심근경색으로, 134 이하이면 건강한 것으로 판별한다. 만약 **cp**가 0 이하라면 **exang**을 확인하고 0 초과면 건강한 것으로 판별하고, **exang**가 0 이하라면 **thal**을 확인해 2 초과면 건강한 것으로, 2 이하면 심근경색으로 판별한다. 두 모델 모두 공통적으로 **cp**, **thal**, **thalach**, **exang**를 활용해 분류를 하며 Pre-Pruned 트리는 **sex** 변수를 추가적으로 활용한다.

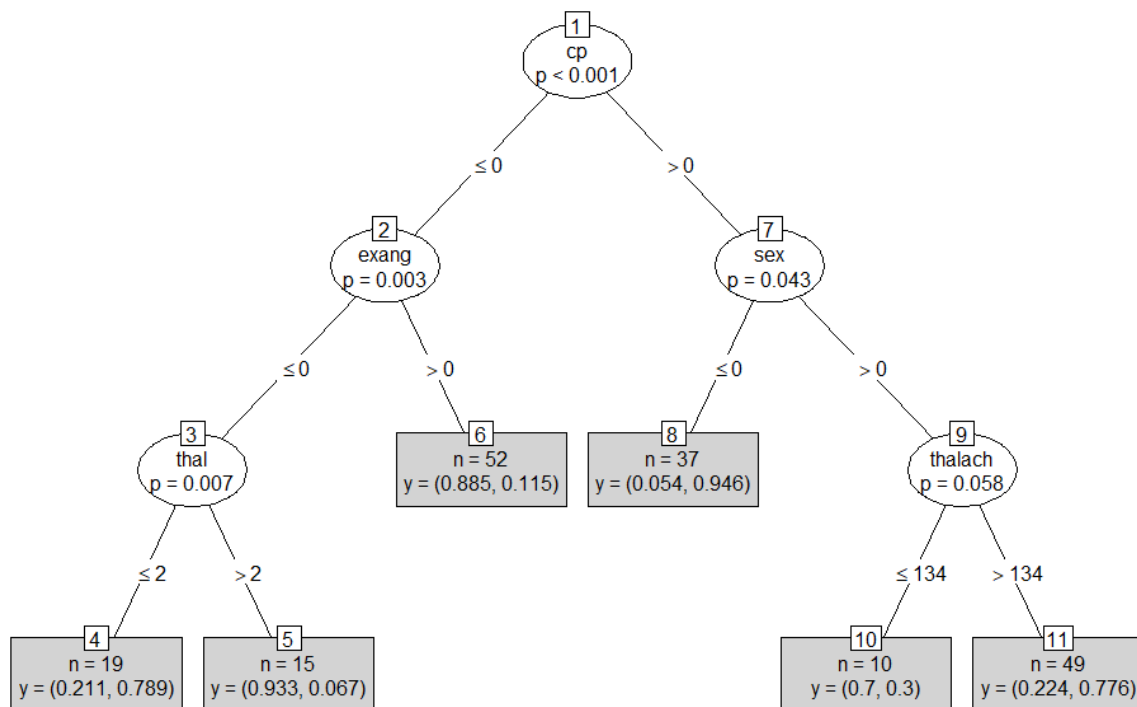


Fig 3. Pre-Pruning 된 트리, (Post-Pruning 결과물은 Fig 2. 참고)

	TPR	Precision	TNR	Accuracy	BCR	F1-measure	Rank
Full Tree	0.8056	0.7826	0.7222	0.7686(2)	0.7630(1)	0.7941(3)	2.00
Post-Pruned	0.8955	0.7500	0.6296	0.7769(1)	0.7509(2)	0.8163(1)	1.33
Pre-Pruned	0.8806	0.7375	0.6111	0.7603(3)	0.7336(3)	0.8027(2)	2.67

Table 5. 이진분류기의 주요 성능지표 비교, 평균 등수

Table 5에서 Full Tree, Post-Pruned, Pre-Pruned Tree의 테스트 데이터에 대한 성능을 비교해볼 수 있다. Pre-Pruned와 Post-Pruned 모델이 Full Tree에 비해 상대적으로 우월한 것을 확인할 수 있는데, 약간의 차이로 Post-Pruned 모델이 가장 우월하다. 이 최적의 모델의 대표적인 규칙 3가지는 다음과 같이 정리할 수 있다. (Fig 2. 참고)

1. **Cp**가 0.5 이상일 경우 심근경색 고위험군이다
2. **Cp**가 0.5 미만, **thal** 이 2.5 이상일 경우 안전하다.
3. **Cp**가 0.5 미만, **thal**이 2.5 미만일 경우 **exang**이 0.5 미만일 때 심근경색 고위험군이다.

Q6

Neural Net

동일한 데이터셋에 대해 인공신경망을 구축하여 이진분류기를 구현한다. 인공신경망의 하이퍼파라미터인 히든 레이어 수(nH)를 {5, 15, 25} 가중치를 0으로 계산하는 threshold인 Decay를 { 5×10^{-15} , 5×10^{-4} , 5×10^{-1} } 최대 iteration을 {100, 500, 1000}으로 조정하겠다. 각 하이퍼파라미터 조합에 대해 학습 데이터를 통해 모델을 만들고 검증 데이터를 통해 AUROC를 도출해 최적의 하이퍼파라미터를 탐색한다. 이때 가능한 27개의 조합에 대해 $nH = 5$, Decay = 5×10^{-1} , 최대 iteration = 100 일때 AUROC는 최대값인 0.7785을 가진다. (*하이퍼파라미터 조합은 supplementary 참고)

이 하이퍼파라미터를 바탕으로 인공신경망을 학습시키고, 테스트 데이터를 사용해 테스트시 다음과 같은 confusion matrix를 얻을 수 있다.

Confusion Matrix		Predicted	
		0	1
Actual	0	36	17
	1	23	45

Table 6. 최적의 하이퍼파라미터로 학습한 NN의 테스트 데이터에 대한 confusion matrix

Q7

모델 비교

로지스틱 회귀분석을 추가적으로 진행해, Post-Pruned Tree(최적), 최적의 Neural Network 총 세가지 모델을 테스트 데이터에 대해 주요 이진분류 성능 지표들(BCR, accuracy, F1-measure)등을 통해 비교하도록 하겠다.

	Estimate	std.error	z-value	Pr(> Z)	
(Intercept)	-0.41447	0.36507	-1.135	0.256240	
age	0.3708	0.38003	0.976	0.329212	
sex	-0.90521	0.38593	-2.346	0.019000	*
cp	1.12692	0.36225	3.111	0.001865	**
trestbps	-0.48934	0.35844	-1.365	0.172193	
chol	-0.63165	0.38967	-1.621	0.105018	
fbs	-0.01308	0.35327	-0.037	0.970463	
restecg	0.20057	0.33425	0.600	0.548466	
thalach	0.93163	0.40629	2.293	0.021845	*
exang	-0.54801	0.39226	-1.397	0.162396	
oldpeak	-0.39889	0.48428	-0.824	0.410126	
slope	-0.11388	0.3753	-0.303	0.761553	
ca	-1.75226	0.52674	-3.327	0.000879	***
thal	-0.40618	0.27925	-1.455	0.145806	

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Confusion Matrix		Predicted	
		0	1
Actual	0	42	11
	1	14	54

Table 7. 로지스틱 회귀분석 모델(상)과 테스트 데이터에 대한 confusion matrix(하)

	TPR	Precision	TNR	Accuracy	BCR	F1-measure	Rank
CART	0.8442	0.8228	0.6818	0.7851(3)	0.7587(3)	0.8333(1)	2.33
NN	0.7973	0.8677	0.8085	0.8017(1)	0.8029(1)	0.8310(3)	1.67
LogisticR	0.7941	0.8308	0.7925	0.7934(2)	0.7933(2)	0.8120(2)	2

Table 8. CART, NN, Logistic Regression 모델의 테스트 데이터에 대한 주요 성능지표, 평균 등수

Table 8의 주요 지표와 Table 3, Table 6, Table 7의 confusion matrix를 확인한다면 인공지능 경향이 전체적으로 가장 성능이 우수하고, 반대로 CART의 성능이 가장 열등한 것을 확인할 수 있다. 이는 모델의 특성을 감안하면 일반적으로 모델의 설명력과 성능이 trade-off 관계에 있기 때문에 의사결정나무의 경우에 성능이 떨어지는 것이 직관과 일치한다. 비록 성능이 약간 낮을 수 있지만, 그 설명력 때문에 본 데이터같이 모델의 설명력이 더 중요한 경우에는 CART를 사용하는 것이 더 좋은 선택일 것이다.

Q8

데이터셋 #2 설정

이미지 인식과 같은 주제는 모델의 설명력보다도 성능이 중요한 대표적인 분야이다. 본 데이터셋은 input 변수로 사람의 얼굴의 주요 지표들, label로 해당 인물의 성별을 사용하는 데이터셋으로, 어느 변수가 주요하게 작용했는지 아는 것보다 남자/여자로 정확하게 판별하는 것이 중요한 사례이다. 본 보고서는 얼굴의 특징을 바탕으로 성별을 예측하는 모델을 구축하겠다. 데이터셋의 크기가 #1보다 크기 때문에 (학습):(검증):(테스트) 데이터를 6:2:2로 분할해 모델을 구축한다.

Feature	Description
long_hair	Whether the person has long hair or not
forehead_width_cm	Width of the person's forehead
forehead_height_cm	Height of the person's forehead
nose_wide	Whether the person has wide nose
nose_long	Whether the person has long nose
lip_thin	Whether the person has thin lips
distance_nose_to_lip_long	How long the distance between lip to nose is
gender	Gender (lable)

출처: <https://www.kaggle.com/datasets/elakiricoder/gender-classification-dataset>

Table 9. 데이터셋 #2의 주요 변수 설명 (N=5001)

Q9

Modeling

Pre-Pruned Tree

데이터#2에 비해 관측치 수가 증가했기 때문에, Q3의 하이퍼파라미터에 추가적으로 **min.split**과 **min.depth**를 더 큰 범위로 탐색하겠다. **min.criterion** = {0.9, 0.95, 0.99, 0.995}, **min.split** = {5, 10, 30, 50}, **max.depth** = {0, 3,4,5,10,20}. 이때 총 96개의 조합에 대해 검증 데이터의 AUROC를 기준으로 최적의 조합 **min.criterion** = 0.9, **min.split** = 5, **max.depth** = 10 이며, 이때 AUROC = 0.9893 이다. 이후 해당 모델을 테스트 데이터에 대해 성능을 평가한다. (*하이퍼파라미터 조합은 supplementary 참고)

Confusion Matrix		Predicted	
		0	1
Actual	0	467	12
	1	25	496

Table 10. 테스트 데이터에 대한 Pre-Pruned Tree의 confusion matrix

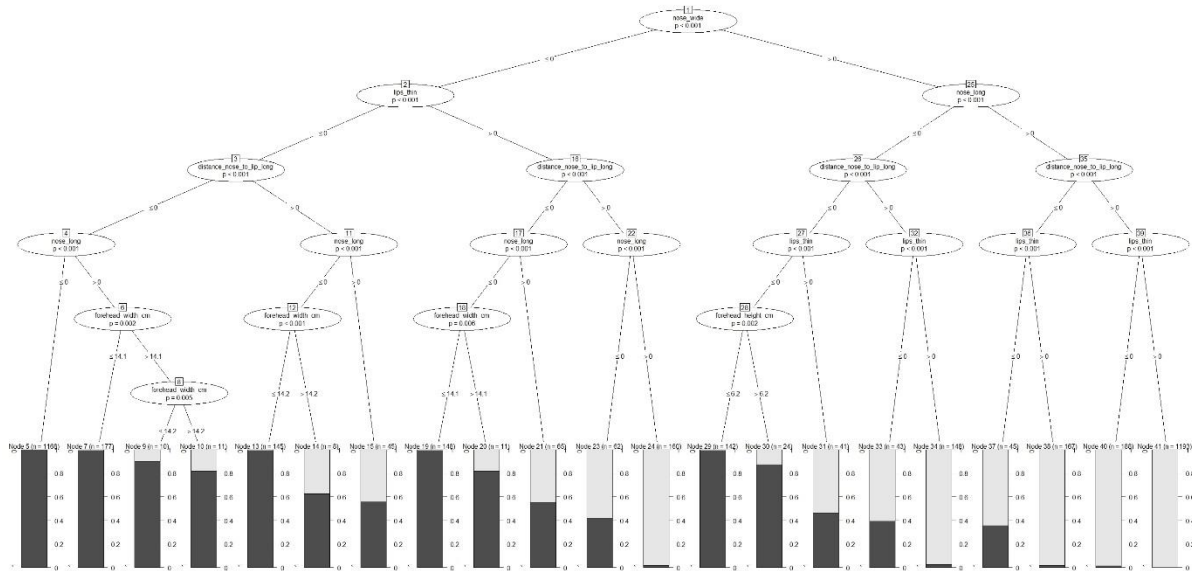


Fig 4. Pre-Pruned Tree 도식화

Neural Network

다음과 같은 하이퍼파라미터 조합을 검증데이터셋에 대한 AUROC를 바탕으로 비교해 최적의 조합을 탐색한다. $nH = \{5, 15, 25\}$, $Decay = \{5 \times 10^{-15}, 5 \times 10^{-4}, 5 \times 10^{-1}\}$, $MaxIt = \{100, 500, 1000\}$. 이때, $nH = 5$, $Decay = 5 \times 10^{-4}$, $Maxit = 1000$ 이 최적의 조합으로, $AUROC = 0.9691220$ 이다. 이후 최적의 인공신경망 모델을 테스트 데이터에 대해 성능을 평가한다. (*하이퍼파라미터 조합은 supplementary 참고)

Confusion Matrix		Predicted	
		0	1
Actual	0	470	17
	1	9	504

Table 11. 테스트 데이터에 대한 Neural Network의 confusion matrix

Q10

모델 비교

Logistic Regression

위 모델들에 추가로 로지스틱 회귀분석을 진행해서 성능을 비교하겠다.

	Estimate	std.error	z-value	Pr(> Z)	
(Intercept)	-0.1508	0.1074	-1.405	0.1600	
long_hair	0.1032	0.1048	0.985	0.3250	
forehead_width_cm	-0.8285	0.1211	-6.839	7.99X10 ⁻¹²	***
forehead_height_cm	-0.4955	0.1082	-4.578	4.69X10 ⁻⁰⁶	***
nose_wide	-2.0651	0.1287	-16.052	<2.00 X10 ⁻¹⁶	***
nose_long	-1.8873	0.1250	-15.098	<2.00 X10 ⁻¹⁶	***
lips_thin	-1.8711	0.1231	-15.198	<2.00 X10 ⁻¹⁶	***
distance_nose_to_lip_long	-1.9493	0.1251	-15.584	<2.00 X10 ⁻¹⁶	***

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Confusion Matrix		Predicted	
		0	1
Actual	0	472	7
	1	15	506

Table 12. Logistic Regression model(상)과 테스트 데이터에 대한 confusion matrix(하)

	TPR	Precision	TNR	Accuracy	BCR	F1-measure	Rank
CART	0.95202	0.97638	0.97495	0.96300(2)	0.96341(2)	0.96404(2)	2
NN	0.97250	0.95010	0.94705	0.96000(3)	0.95969(3)	0.96117(3)	3
LogisticR	0.97121	0.98635	0.98539	0.97800(1)	0.97827(1)	0.97872(1)	1

Table 13. 세가지 모델의 테스트 데이터에 대한 이진분류기 성능지표 비교

직관과 상반되게 인공신경망이 CART에 비해 우수하지 못하며, 로지스틱 회귀분석이 가장 우수한 것으로 나타났다. 이는 두가지 방식으로 해석이 가능하다. 우선 인공신경망 구축 과정에서 충분한 범위의 하이퍼파라미터를 탐색하지 못해, 충분히 좋은 모델을 구축하는데 실패했기 때문에 비교적 하이퍼파라미터가 단순한 CART와 필요없는 로지스틱 회귀분석이 더 우수하게 나타난 것이다. 두번째로, 인공신경망은 그 특성상 많은 데이터 양을 필요로 하고, 복잡한 관계를 설명하는데 적합한 모델이지만, 본 데이터셋이 그리 크지 않고, 변수의 개수가 적기 때문에 단순한 모델이 더 적합했다고 해석할 수 있다. 이때 로지스틱 회귀분석이 퍼셉트론 1개인 일종의 가장 단순한 인공신경망이며, 본 보고서에서 하이퍼파라미터 탐색 과정에서 최소 5개 이상의 레이어를 쌓고 탐색했다는 것을 고려한다면 데이터셋에 내제된 차원이 작아 단순한 모델이 더 적합했다는 설 힘이 실린다.

Supplementary

1. Data#1 CART Pre-Pruning Hyperparameters, arranged by AUROC

min.criterion	min.split	max.depth	0.733	BCR	F1	AUROC
0.9	10	0	0.733	0.741	0.75	0.815
0.9	10	10	0.733	0.741	0.75	0.815
0.9	10	5	0.733	0.741	0.75	0.815
0.9	30	0	0.733	0.741	0.75	0.815
0.9	30	10	0.733	0.741	0.75	0.815
0.9	30	5	0.733	0.741	0.75	0.815
0.9	50	0	0.733	0.741	0.75	0.815
0.9	50	10	0.733	0.741	0.75	0.815
0.9	50	5	0.733	0.741	0.75	0.815
0.95	10	0	0.733	0.741	0.75	0.815
0.95	10	10	0.733	0.741	0.75	0.815
0.95	10	5	0.733	0.741	0.75	0.815
0.95	30	0	0.733	0.741	0.75	0.815
0.95	30	10	0.733	0.741	0.75	0.815
0.95	30	5	0.733	0.741	0.75	0.815
0.95	50	0	0.733	0.741	0.75	0.815
0.95	50	10	0.733	0.741	0.75	0.815
0.95	50	5	0.717	0.741	0.75	0.815
0.99	10	0	0.717	0.722	0.691	0.783
0.99	10	10	0.717	0.722	0.691	0.783
0.99	10	5	0.717	0.722	0.691	0.783
0.99	30	0	0.717	0.722	0.691	0.783
0.99	30	10	0.717	0.722	0.691	0.783
0.99	30	5	0.717	0.722	0.691	0.783
0.99	50	0	0.717	0.722	0.691	0.783
0.99	50	10	0.717	0.722	0.691	0.783
0.99	50	5	0.733	0.722	0.691	0.783
0.9	100	0	0.733	0.741	0.75	0.743
0.9	100	10	0.733	0.741	0.75	0.743
0.9	100	5	0.733	0.741	0.75	0.743
0.95	100	0	0.733	0.741	0.75	0.743
0.95	100	10	0.733	0.741	0.75	0.743
0.95	100	5	0.733	0.741	0.75	0.743
0.99	100	0	0.733	0.741	0.75	0.743
0.99	100	10	0.733	0.741	0.75	0.743
0.99	100	5	0.733	0.741	0.75	0.743

2. Data #1 Neural Network Hyperparameters, arranged by AUROC

nH	Decay	MaxIt	AUROC
5	0.5	100	0.778
15	0.5	500	0.770
25	0.5	100	0.770
25	0.5	500	0.770
25	0.5	1000	0.770
5	0.5	500	0.762
5	0.5	1000	0.762
15	0.5	100	0.762
15	0.5	1000	0.762
15	5.00E-10	100	0.754
25	5.00E-10	500	0.754
15	5.00E-04	100	0.746
25	5.00E-10	100	0.746
15	5.00E-10	1000	0.730
5	5.00E-10	100	0.713
5	5.00E-04	1000	0.713
25	5.00E-04	100	0.698
15	5.00E-10	500	0.697
5	5.00E-10	1000	0.690
15	5.00E-04	500	0.689
5	5.00E-04	100	0.688
15	5.00E-04	1000	0.681
5	5.00E-04	500	0.673
25	5.00E-04	1000	0.673
5	5.00E-10	500	0.673
25	5.00E-04	500	0.664
25	5.00E-10	1000	0.602

3. Data #2 CART Pre-Pruning Hyperparameters, arranged by AUROC

min.criterion	min.split	max.depth	ACC	BCR	F1	AUROC
0.9	5	0	0.959	0.958	0.957	0.959
0.9	5	5	0.959	0.958	0.957	0.959
0.9	5	10	0.959	0.958	0.957	0.959
0.9	5	20	0.959	0.958	0.957	0.959
0.9	10	0	0.959	0.958	0.957	0.959
0.9	10	5	0.959	0.958	0.957	0.959
0.9	10	10	0.959	0.958	0.957	0.959

0.9	10	20	0.959	0.958	0.957	0.959
0.9	30	0	0.959	0.958	0.957	0.959
0.9	30	5	0.959	0.958	0.957	0.959
0.9	30	10	0.959	0.958	0.957	0.959
0.9	30	20	0.959	0.958	0.957	0.959
0.9	50	0	0.959	0.958	0.957	0.959
0.9	50	5	0.959	0.958	0.957	0.959
0.9	50	10	0.959	0.958	0.957	0.959
0.9	50	20	0.959	0.958	0.957	0.959
0.95	5	0	0.959	0.958	0.957	0.959
0.95	5	5	0.959	0.958	0.957	0.959
0.95	5	10	0.959	0.958	0.957	0.959
0.95	5	20	0.959	0.958	0.957	0.959
0.95	10	0	0.959	0.958	0.957	0.959
0.95	10	5	0.959	0.958	0.957	0.959
0.95	10	10	0.959	0.958	0.957	0.959
0.95	10	20	0.959	0.958	0.957	0.959
0.95	30	0	0.959	0.958	0.957	0.959
0.95	30	5	0.959	0.958	0.957	0.959
0.95	30	10	0.959	0.958	0.957	0.959
0.95	30	20	0.959	0.958	0.957	0.959
0.95	50	0	0.959	0.958	0.957	0.959
0.95	50	5	0.959	0.958	0.957	0.959
0.95	50	10	0.959	0.958	0.957	0.959
0.95	50	20	0.959	0.958	0.957	0.959
0.99	5	0	0.959	0.958	0.957	0.959
0.99	5	5	0.959	0.958	0.957	0.959
0.99	5	10	0.959	0.958	0.957	0.959
0.99	5	20	0.959	0.958	0.957	0.959
0.99	10	0	0.959	0.958	0.957	0.959
0.99	10	5	0.959	0.958	0.957	0.959
0.99	10	10	0.959	0.958	0.957	0.959
0.99	10	20	0.959	0.958	0.957	0.959
0.99	30	0	0.959	0.958	0.957	0.959
0.99	30	5	0.959	0.958	0.957	0.959
0.99	30	10	0.959	0.958	0.957	0.959
0.99	30	20	0.959	0.958	0.957	0.959
0.99	50	0	0.959	0.958	0.957	0.959
0.99	50	5	0.959	0.958	0.957	0.959
0.99	50	10	0.959	0.958	0.957	0.959

0.99	50	20	0.959	0.958	0.957	0.959
0.99	5	0	0.959	0.958	0.957	0.959
0.99	5	5	0.959	0.958	0.957	0.959
0.99	5	10	0.959	0.958	0.957	0.959
0.99	5	20	0.959	0.958	0.957	0.959
0.99	10	0	0.959	0.958	0.957	0.959
0.99	10	5	0.959	0.958	0.957	0.959
0.99	10	10	0.959	0.958	0.957	0.959
0.99	10	20	0.959	0.958	0.957	0.959
0.99	30	0	0.959	0.958	0.957	0.959
0.99	30	5	0.959	0.958	0.957	0.959
0.99	30	10	0.959	0.958	0.957	0.959
0.99	30	20	0.959	0.958	0.957	0.959
0.99	50	0	0.959	0.958	0.957	0.959
0.99	50	5	0.959	0.958	0.957	0.959
0.99	50	10	0.959	0.958	0.957	0.959
0.99	50	20	0.959	0.958	0.957	0.959
0.9	5	4	0.959	0.958	0.957	0.959
0.9	10	4	0.959	0.958	0.957	0.959
0.9	30	4	0.959	0.958	0.957	0.959
0.9	50	4	0.959	0.958	0.957	0.959
0.95	5	4	0.959	0.958	0.957	0.959
0.95	10	4	0.959	0.958	0.957	0.959
0.95	30	4	0.959	0.958	0.957	0.959
0.95	50	4	0.959	0.958	0.957	0.959
0.99	5	4	0.959	0.958	0.957	0.959
0.99	10	4	0.959	0.958	0.957	0.959
0.99	30	4	0.959	0.958	0.957	0.959
0.99	50	4	0.959	0.958	0.957	0.959
0.99	5	4	0.959	0.958	0.957	0.959
0.99	10	4	0.959	0.958	0.957	0.959
0.99	30	4	0.959	0.958	0.957	0.959
0.99	50	4	0.959	0.958	0.957	0.959
0.9	5	3	0.96	0.96	0.958	0.96
0.9	10	3	0.96	0.96	0.958	0.96
0.9	30	3	0.96	0.96	0.958	0.96
0.9	50	3	0.96	0.96	0.958	0.96
0.95	5	3	0.96	0.96	0.958	0.96
0.95	10	3	0.96	0.96	0.958	0.96
0.95	30	3	0.96	0.96	0.958	0.96

0.95	50	3	0.96	0.96	0.958	0.96
0.99	5	3	0.96	0.96	0.958	0.96
0.99	10	3	0.96	0.96	0.958	0.96
0.99	30	3	0.96	0.96	0.958	0.96
0.99	50	3	0.96	0.96	0.958	0.96
0.99	5	3	0.96	0.96	0.958	0.96
0.99	10	3	0.96	0.96	0.958	0.96
0.99	30	3	0.96	0.96	0.958	0.96
0.99	50	3	0.96	0.96	0.958	0.96

4. Data #2 Neural Net Hyperparameters, arranged by AUROC

nH	Decay	Maxit	AUROC
25	5.00E-04	1000	0.966
15	5.00E-10	500	0.966
25	5.00E-04	500	0.965
25	5.00E-20	500	0.964
15	5.00E-20	500	0.963
15	5.00E-10	1000	0.963
15	5.00E-20	100	0.963
5	5.00E-20	500	0.963
15	5.00E-04	100	0.963
25	5.00E-10	500	0.962
5	5.00E-04	500	0.962
5	5.00E-10	100	0.962
25	5.00E-10	1000	0.962
5	0.5	500	0.962
5	0.5	1000	0.961
15	0.5	100	0.961
15	0.5	500	0.961
15	0.5	1000	0.961
25	0.5	500	0.961
15	5.00E-04	500	0.961
5	5.00E-04	1000	0.961
5	0.5	100	0.960
25	0.5	100	0.960
25	0.5	1000	0.960
5	5.00E-20	1000	0.960
5	5.00E-10	1000	0.960
5	5.00E-10	500	0.960

15	5.00E-04	1000	0.960
25	5.00E-20	1000	0.960
15	5.00E-20	1000	0.960
15	5.00E-10	100	0.960
5	5.00E-04	100	0.959
25	5.00E-20	100	0.957
25	5.00E-10	100	0.929
25	5.00E-04	100	0.887
5	5.00E-20	100	0.877