

класифікація процесів (дискретні, неперервні, випадкові, керовані та інші).[1]

Математичне моделювання діяльності підприємств повинне бути засноване на аналізі діяльності підприємства й, у свою чергу, збагачувати аналіз результатами й висновками, отриманими після рішення відповідних задач. У суспільстві з ринковою економікою застосування математичних моделей дає певний ефект на мікроекономічному рівні.[2] Економічний стан підприємства передбачає дослідження усіх аспектів його діяльності. Проте, у багатьох науковців переважає дослідження фінансової складової. Звісно, вона є дуже ваговою, проте вважаємо за недоцільне нівелювання інших сфер діяльності підприємства.

Моделі різного виду систем, функціонування котрих залежить від ряду випадкових факторів, можуть бути сформульованими з допомогою термінів, так званих випадкових процесів. Для їх аналізу будуються імовірнісні моделі поведінки, що враховують вплив цих випадкових факторів на значення основних шуканих параметрів і оцінюють рівень ризику.

Список літератури

1. Куліков П. М., Іващенко Г. А. Економіко-математичне моделювання фінансового стану підприємства: навч. посібн. Х.: ІНЖЕК, 2009. 178 с.
2. Агапова А. Економіко-математичне моделювання оцінки економічного стану підприємства з використанням шкали бажаності Харрінгтона. Економічний аналіз. Вип. 8. 2011. С. 15-18.

УДК 004.8

*Хмелівський Ю.С., студент 1 курсу спеціальності «Комп'ютерні технології обробки даних (DataScience)»
Нескородева Т.В., д.т.н., доцент, завідувач кафедри інформаційних технологій*

ПОБУДОВА НЕЛІНІЙНОГО КЛАСИФІКАТОРА ЗАХВОРЮВАННЯ СЕРЦЯ

Донецький національний університет імені Василя Стуса, м. Вінниця

Лінійні моделі відносно просто описати та побудувати і вони мають перевагу порівняно з іншими підходами завдяки інтерпретованості та можливості зробити статистичні висновки. Однак у стандартної лінійної регресії можуть бути суттєві недоліки, пов'язані з її передбачувальною силою. Це зумовлено тим, що припущення про лінійність майже завжди є наближенням, а іноді дуже поганим наближенням. Ми можемо покращити результати, одержувані за допомогою методу найменших квадратів, застосовуючи гребеневу регресію, метод Лассо та інші підходи. покращення при використанні цих

методів досягається шляхом спрощення лінійної моделі, а отже, завдяки зниженню дисперсії оцінок коефіцієнтів. Однак ми все ще використовуємо лінійну модель, яку далі не покращити![1]

Існує підхід при якому потрібно послабити припущення про лінійність, намагаючись зберегти модель настільки інтерпретованою, наскільки це можливо. Існує кілька розширень лінійної моделі, такі як поліноміальна регресія, ступінчасті функції, сплайни, локальна агресія та узагальнені адитивні моделі.

У цій роботі розглянемо узагальнені адитивні моделі чи GAM. GAM (generalized additive model) розширюють стандартну лінійну модель, допускаючи використання нелінійних функцій кожного предиктора при одночасному збереженні адитивності. Подібно до лінійної моделі GAM можна застосовувати як для кількісних так і для якісних відгуків.

Природним способом розширення множинної лінійної регресійної моделі для вираження нелінійних залежностей між кожною змінною та відгуком є заміна кожного лінійного компонента $\beta_j x_{ij}$ на нелінійну функцію $f_j(x_{ij})$. Тоді ми могли б записати модель так:

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \dots + f_p(x_{ip}) + \varepsilon_i.$$

Це-приклад GAM. Ця модель називається адитивною, оскільки ми обчислюємо окремі функції для кожної змінної, а потім підсумовуємо їх.

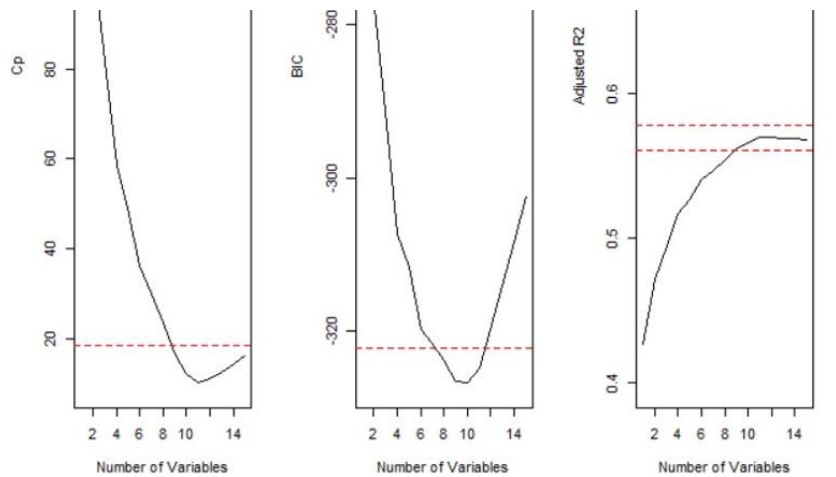
Розглянемо метод на прикладі датасета створеного для визначення захворювання серця.[2] Він містить 918 унікальних рядків та 11 стовпців із характеристиками людини і один бінарний стовпець відгук.

Атрибути:

1. Age: вік пацієнта [років]
2. Sex: стать пацієнта [М: Чоловік, Ж: Жінка]
3. ChestPainType: тип болю в грудях [ТА: Типова стенокардія, АТА: Атипична стенокардія, NAP: неангінальний біль, ASY: безсимптомний]
4. RestingBP: артеріальний тиск у стані спокою [мм рт.ст.]
5. Cholesterol: сироватковий холестерин [мм/дл]
6. FastingBS: рівень цукру в крові натще [1: якщо FastingBS > 120 мг/дл, 0: інакше]
7. RestingECG: результати електрокардіограми в спокої [Normal: Нормальний, ST: наявність аномалій зубця ST-T (інверсії зубця Т та/або елевація або депресія ST > 0,05 мВ), LVH: ймовірна або визначена гіпертрофія лівого шлуночка за критеріями Естеса]
8. MaxHR: досягнута максимальна частота серцевих скорочень [числове значення від 60 до 202]
9. ExerciseAngina: стенокардія, спричинена фізичним навантаженням [Y: Так, N: Ні]
10. Oldpeak: oldpeak = ST [Числове значення, виміряне в депресії]
11. ST_Slope: нахил сегмента ST піку вправи на кардіограмі [Up: вгору, Flat: рівна, Down: вниз]
12. HeartDisease: вихідний клас [1: захворювання серця, 0: нормально].

В нас є 5 кількісних предикторів, тому потрібно їх перевести в числове значення [3].

Розбиваємо данні на навчальну і тестову вибірки. HeartDisease як відгук, решта предиктори. Знаходимо модель в яку входять лише деякі з предикторів, виконавши відбір поступовим виключенням.

Рисунок 7 - Оцінки cp , BIC та Adjusted R2

Усі оцінки CP , BIC та $adjr2$ показують, що розмір 8 є мінімальним розміром для підмножини, для якої оцінки знаходяться в межах 0,2 стандартних відхилень від оптимального. Ми вибираємо 8 як найкращий розмір підмножини та знаходимо найкращі 8 змінних: Cholesterol, FastingBS, Oldpeak, cptASY, sex, eaYes, stFlat та stUp.

Оскільки, змінна Cholesterol єдина з вище вказаних не є бінарною, то нелінійний зв'язок можливий тільки для неї. Побудуємо GAM.

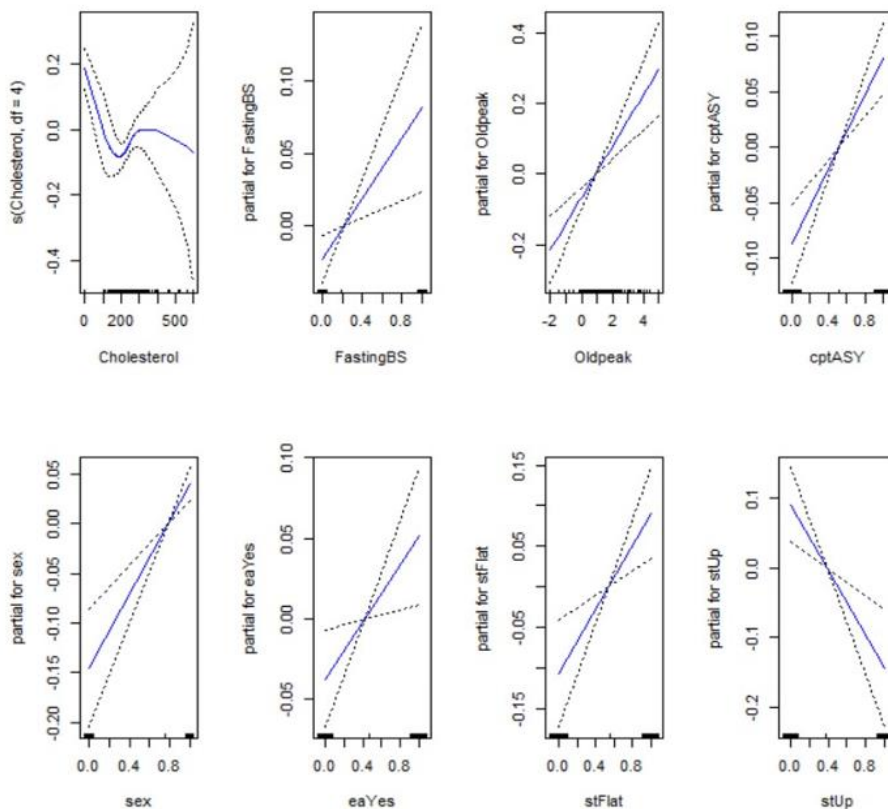


Рисунок 8 – Графіки GAM

Ми отримуємо тест R-квадрат 0,556 за допомогою GAM з 8 предикторами.

```
Anova for Nonparametric Effects
      Npar Df Npar F      Pr(F)
(Intercept)
s(Cholesterol, df = 4)      3 7.2552 9.225e-05 ***
```

Непараметричний тест Anova показує переконливі докази нелінійного зв'язку між відгуком та Cholesterol.

Список літератури

1. Джеймс Г., Уиттон А., Хасті Т., Тибиширани Р. Введение в статистическое обучение с примерами на языке R Изд. Второе, испр. Пер с англ. С.Э. Мастицкого –М. ДМК Пресс, 2017. -456с
2. Heart Failure Prediction Dataset [Електронний ресурс] . – Режим доступу до ресурсу: <https://www.kaggle.com/fedesoriano/heart-failure-prediction>
3. Хмелівський Ю.С., Нескородева Т.В. Аналіз даних для прогнозування серцевої недостатності засобами мови R. Матеріали II всеукраїнської науково-практичної конференції для студентів, аспірантів та молодих вчених "Комп'ютерні технології обробки даних" (10 грудня 2021 року) - Вінниця: ДонНУ імені Василя Стуса., с.57-60.
4. Резнік Р.Ю., Нескородева Т.В. Класифікація зразків скла на основі хіміко-фізичних властивостей методами статистичного навчання. Матеріали II всеукраїнської науково-практичної конференції для студентів, аспірантів та молодих вчених "Комп'ютерні технології обробки даних" (10 грудня 2021 року) - Вінниця: ДонНУ імені Василя Стуса., с.52-55.

УДК 004.8

Швець Х.І., студентка 2 курсу спеціальності 122 «Комп'ютерні науки»

Потапова Н. А., к.е.н., доцент, доцент кафедри інформаційних технологій

УТОЧНЕННЯ НАБЛИЖЕНИХ РОЗВ'ЯЗКІВ МЕТОДОМ ПРОСТОЇ ІТЕРАЦІЇ

Донецький національний університет імені Василя Стуса, м. Вінниця

Ізолювавши інтервал, на якому існує один корінь, необхідно вибрати конкретний алгоритм знаходження кореня із заданою точністю. Алгоритми уточнення коренів поділяються на дві категорії – алгоритми звуження інтервалу та ітераційні алгоритми. Вибір алгоритму для чисельного знаходження кореня проводиться з урахуванням його ефективності. Алгоритм повинен проводити якомога менше обчислень функції, тобто працювати швидко, але, водночас, бути простим при програмуванні й застосуванні. Ітераційні алгоритми потребують перевірки на збіжність. Існує також велика кількість різноманітних комбінованих методів.

Метод простої ітерації застосовують для розв'язування задач про нерухому точку, тобто рівнянь вигляду: $x=f(x)$. В цьому методі процес розв'язання