

# Огляд економічних моделей на основі методів машинного навчання для прогнозування рецесій

Козюпа Тарас Костянтинівич, магістрант  
ORCID ID 0000-0002-8061-9430

Науковий керівник: Шевчук Олена Анатоліївна, д.е.н., проф.  
КПІ ім. Ігоря Сікорського, Україна

З плином часу, велика увага була зосереджена на оцінці економічних циклів і визначенні їх основних поворотних (критичних) моментів. Природа та причини рецесій неоднакові; тобто рецесії не спричиняються тими самими економічними потрясіннями (шоками). Тому на даному етапі розвитку технологій широке застосування знаходять різноманітні методи машинного навчання для прогнозування рецесій. Зокрема, методи машинного навчання включають методи регуляризації, такі як *Ridge*, оператор найменшого абсолютного скорочення та вибору (*Least Absolute Shrinkage and Selection Operator, LASSO*), еластична мережа (*Elastic Net*), класифікатори дискримінантного аналізу, байєсівські (*Bayesian*) класифікатори та дерева класифікації та регресії (*Classification and regression trees, CART*), такі як *Bagging, Random Forests, Boosting* тощо. Останнім часом використовуються більш спеціалізовані індикатори, такі як індекс випереджаючих кредитів (*Leading Credit Index, LCI*), який відображає потенційні структурні зміни на фінансових ринках [1].

Бінарні регресійні моделі типу *Logit* із вибраваною імовірністю (*Penalized likelihood*). Модель *Ridge Logit* (на основі моделі регресії *Ridge* [2]) — це підхід до моделювання, який функціонує шляхом регуляризації або скорочення оціночних коефіцієнтів до нуля. Даний процес має низку переваг, оскільки він призводить до меншої дисперсії, зменшує складність базової моделі та не зменшує кількість предикторів, а лише послаблює їхній ефект.

Альтернативним підходом є модель найменшого абсолютного скорочення та оператора вибору (*LASSO*), представлена в [3]. Встановлюючи інший тип скорочення, модель *LASSO Logit* забезпечує кращу інтерпретацію та може виконувати вибір змінних для визначення предикторів, які тісно пов'язані з рецесіями. Метод еластична мережа (*Elastic Net*) також використовується в економічному аналізі. *Elastic Net* — це альтернативний підхід до моделювання, який було розширено до системи регресії *Logit*. Еластична мережа була представлена як вдосконалена технологія (порівняно з *LASSO* та *Ridge*), яка здатна обробляти висококорельовані змінні в наборі предикторів і враховувати проблему колінеарності, притаманну аналізу даних великих об'ємів. Він зберігає основні переваги та виключає недоліки двох попередніх методів.

Методи дискримінантного аналізу використовуються для класифікації, перебування економіки в рецесії в певний період часу. Зокрема, при аналізі застосовується лінійний і регуляризований дискримінантний аналіз.

Лінійний дискримінантний аналіз (*Linear Discriminant Analysis, LDA*) — це багатозмінний статистичний метод, який можна використовувати для класифікації та прогнозування поведінки бінарної залежної змінної на основі набору предикторів. Основна ідея полягає в тому, щоб вивести лінійну комбінацію предикторів, яка «найкраще» диференціює події, пов'язані з конкретним станом (рецесія або розвиток економіки).

Алгоритми рекурсивного розподілення (*Recursive Partitioning, rpart*) використовуються для побудови моделей класифікації та регресії; тобто отримані моделі можна представити у вигляді бінарних дерев. Мета аналізу полягає в тому, щоб передбачити/класифікувати, чи перебуває економіка в рецесії, на основі набору прогностичних змінних. Деревоподібна модель будується за допомогою двоетапної процедури. На першому кроці виявляється єдина змінна предиктора, яка найкраще розбиває дані на дві групи. Другий етап процедури полягає в обрізанні (формуванні) всього дерева.

*Bootstrap aggregating* або *Bagging* [4] (пакування, від англ. 'bag' – мішок, пакунок) є широко застосовуваним ансамблевим методом (тобто таким, що застосовується в сфері машинного навчання) класифікації. Він покращує точність прогнозування класифікатора,

генеруючи кілька версій класифікатора на основі наданих копій набору даних, а потім об'єднує ці класифікатори для створення єдиного. Спеціальна процедура створює стабільний прогноз/класифікатор з меншою дисперсією і, таким чином, досягаються значні покращення по відношенню до точності. Для опису алгоритму пакетування, розглянемо початкову вибірку,

$$(Y_t, X_{t-h}), t = 1, \dots, T$$

де  $Y_t$  є бінарною залежною змінною рецесії в момент часу  $t$ , а  $X_{t-h}$  є вектором предикторів у момент часу  $t - h$ .

В аналізі (проблеми класифікації рецесії) нас цікавить класифікація того, чи перебуває економіка в рецесії. Іншими словами, ми зацікавлені в оцінці

$$\text{Prob}(Y_t = 1 | \mathcal{R}_{t-h}), \text{ або } \text{Prob}(Y_t = 0 | \mathcal{R}_{t-h}),$$

тобто умовної ймовірності рецесії в момент часу  $t$ , враховуючи набір інформації про предиктори в момент часу  $t - h$ ,  $\mathcal{R}_{t-h}$ .

Основна ідея алгоритму полягає у створенні  $B$  початкових зразків  $S_1, S_2, \dots, S_B$  вибірки. З кожної початкової вибірки  $S_i, i = 1, \dots, B$ , величина інтересу (*quantity of interest*), тобто будь-якого класифікатора, скажімо  $C_i$ , оцінюється на основі тієї самої процедури машинного навчання. Тоді пакетний оцінювач/класифікатор (*bagged estimator/classifier*),  $C_{\text{bag}}$ , можна отримати шляхом агрегування різних початкових класифікаторів.

Модель векторної авторегресії (Vector Autoregression Model, VAR). Модель VAR є багатоваріантною формою, розширеною з моделі авторегресії (AR). Модель AR описує, як часовий ряд  $r_t$  може бути спрощений до його *lag* порядків (*lag orders*). Для багатовимірного вектора часового ряду  $X_t$ , де  $a_t$  — вектор білого шуму, VAR ( $p$ ) можна виразити як:

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + a_t.$$

Модель довго- та короткострокової пам'яті (Long-short Term Memory Model, LSTM). Як тип моделі рекурентної нейронної мережі (Recurrent neural network, RNN), LSTM — це метод машинного навчання, який добре обробляє дані часових рядів. Метод LSTM був запропонований як вдосконалений варіант традиційного RNN, оскільки він здатен накопичувати довгострокову пам'ять шляхом додавання блоку пам'яті та використання входів (*gates*) для керування збереженням та очищенням пам'яті.

Три входи моделі LSTM називаються: вхідний, вихідний і очищення.

Таким чином, методи машинного навчання слугують потужним інструментом в області економічного прогнозування, зокрема моделювання прогнозів економічних рецесій. Деякі з них слугують безпосередньо для прогнозів, інші ж — для покращення точності прогнозування.

#### Література:

1. Liu, W., & Moench, E. (2016). What predicts US recessions? *International Journal of Forecasting*, 32(4), 1138–1150. <https://doi.org/10.1016/j.ijforecast.2016.02.007>.
2. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1), 69–82. <https://doi.org/10.1080/00401706.1970.10488635>.
3. Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
4. Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1007/bf00058655>