

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОСТРОЕНИЯ ЛИНЕЙНЫХ МОДЕЛЕЙ ЗАДАЧ СЕЛЬСКОГО ХОЗЯЙСТВА

Сабитов Б.Р.¹, Керимов У.Т.², Ачекеев К.С.³, Мусуралиева Д.Э.⁴, Мустапакулов Н.К.⁵

(1) доцент кафедры «Информационных технологии и программирования» КНУ им. Ж.Баласагына, baratbek@mail.ru

(2) и.о. доцента кафедры «Прикладная информатика» КГУ им. И.Арабаева, u.kerimov@bk.ru

(3) Старший преподаватель кафедры «Прикладная информатика» КГУ им. И.Арабаева, kadyrbekachekeev@mail.ru

(4) Старший преподаватель кафедры «Информационных технологии и программирования» КНУ им. Ж.Баласагына, r.musuralieva@yandex.com

(5) Магистрант КГУ имени И. Арабаева, mustapakulov@mail.ru

Аннотация: В этой статье изучается применение технологии машинного обучения для задач сельского хозяйства. Рассмотрены технологии построения моделей в растениеводстве с применением регрессионного анализа. Изучены два подхода к построению моделей с линейными и нелинейными данными с применением градиентного метода оптимизации.

Ключевые слова: Scikit Learn, АРК, линейные модели, нелинейные модели, градиентный спуск.

АЙЫЛ ЧАРБА КӨЙГӨЙҮНҮН СЫЗЫКТУУ МОДЕЛДЕРИН ТҮЗҮҮ ҮЧҮН МАШИНАЛЫК ОКУТУУНУН АЛГОРИТИМДЕРИН КОЛДОНУУ

Сабитов Б.Р.¹, Керимов У.Т.², Ачекеев К.С.³, Мусуралиева Д.Э.⁴, Мустапакулов Н.К.⁵

(1) Ж. Баласагын атындагы КУУнун «Маалыматтык технологиялар жана программалоо» кафедрасынын доценти, baratbek@mail.ru

(2) И. Арабаев атындагы КМУнун «Колдонмо информатика» кафедрасынын доцентинин м.а., u.kerimov@bk.ru

(3) И. Арабаев атындагы КМУнун «Колдонмо информатика» кафедрасынын ага окутуучусу, kadyrbekachekeev@mail.ru

(4) Ж. Баласагын атындагы КУУнун «Маалыматтык технологиялар жана программалоо» кафедрасынын ага окутуучусу, r.musuralieva@yandex.com

(5) И. Арабаев атындагы КМУнун магистранты, mustapakulov@mail.ru

Аннотациясы: Бул макалада айыл чарба көйгөйлөрүнө машина менен окутуу технологиясынын колдонулушу изилденген. Регрессиялык анализдин жардамы менен өсүмдүктөрдү өстүрүүдө моделдерди түзүү технологиялары каралды. Сызыктуу жана сызыктуу эмес маалыматтар менен моделдерди курууга эки ыкма градиенттик оптималдаштыруу ыкмасынын жардамы менен изилденет.

Өзөктүү сөздөр: Scikit Learn, АПК, Сызыктуу моделдер, Сызыктуу эмес моделдер, Градиент ыкмасы.

USING MACHINE LEARNING ALGORITHMS TO BUILD LINEAR MODELS OF AGRICULTURAL PROBLEMS

Sabitov B.R.¹, Kerimov U.T.², Achekeev K.S.³, Musuralieva D.E.⁴, Mustapakulov N.K.⁵

(1) Associate Professor of the Department of "Information Technology and Programming" KNU. J. Balasagyn, baratbek@mail.ru

(2) acting Associate Professor of the Department of "Applied Informatics" of the KSU named after I. Arabaev, u.kerimov@bk.ru

(3) senior lecturer of the Department of "Applied Informatics" of the KSU named after I. Arabaev, kadyrbekachekeev@mail.ru

(4) senior lecturer of the Department of the "Information Technology and Programming" KNU. J. Balasagyn, r.musuralieva@yandex.com

(5) Master student of KSU named after I. Arabaev, mustapakulov@mail.ru

Abstract: This article explores the application of machine learning technology for agricultural tasks. The technologies for building models in crop production using regression analysis are considered. Two approaches to building models with linear and non-linear data using the gradient optimization method are studied.

Keywords: Scikit Learn, APK, linear models, non-linear models, gradient descent.

В настоящей работе рассматривается технологии применения машинного обучения для прикладных задач. Машинное обучения, это прикладное направление искусственного интеллекта, изучающего интеллектуальные системы и библиотеки Python, способные обучаться на базе данных и делать на этой основе прогнозы. В работе также изучается основные методы прогнозирования, в которых, приводится описание основных направлений, задачи алгоритмов машинного обучения. В частности, изучается задача прогнозирования урожайности сельхоз культур с применением алгоритмов машинного обучения. В работе также использовано великолепная система Scikit Learn для анализа многих прикладных задач. Создадим пакет программ для различных моделей АПК с применением различных технологий машинного обучения. Используя линейные и нелинейные данные построены модели в растениеводстве и животноводстве. Обычно параметры моделей определяется с использованием функции издержек, которое определяется функционалом

$$J(X, h_{\theta}) = \frac{1}{m} \sum_{i=1}^m (\theta^T * x^{(i)} - y^{(i)})^2 \quad (1)$$

Необходимо правильно выбрать алгоритм обучения модели и оптимальное значение гиперпараметров для изучаемой задачи. Изучим два различных технологий обучения моделей.

Первое, применение прямого уравнения в виде нормального уравнения, из которой можно определить параметры модели.

Второй подход применение численного подхода к вычислению параметров модели с применением градиентного метода оптимизации. Метод градиентного спуска постепенно корректирует параметры модели, для нахождения минимума функционала.

Изучим конкретные примеры по построению моделей по реализации продукции в животноводстве и растениеводстве.

Рассматриваются сложные нелинейные данные, которые описывают количество реализованной продукции в отраслях растениеводства и животноводства и построим их модель на основе регрессионного анализа. В частности, рассмотрены сложные нелинейные модели реализации поголовья скота в животноводстве, зерновых и зерново-бобовых в растениеводстве используя различные модели.

Рассмотрим сначала первый подход к построению моделей в АПК. Изучим следующую задачу оптимизации

$$\min_{\theta} J(X, h_{\theta}) = \min_{\theta} \frac{1}{m} \sum_{i=1}^m (\theta^T * x^{(i)} - y^{(i)})^2 \quad (2)$$

Параметры модели θ можно определить с помощью прямого уравнения в виде нормального уравнения. Найденное значения θ , которое сводит к минимуму функцию издержек, записывается в аналитическом виде иными словами, математическое уравнение, дающее результат напрямую в виде формулы. Она называется нормальным уравнением и представляется в виде решения

$$\hat{\theta} = (X^T * X)^{-1} X^T y \quad (3)$$

К сожалению, инвертирование, т.е. найти обратную матрицу , матрицы $X^T X$ при нечеткой мультиколлинеарности численно нестабильно, но существует решение.

Известно также, что полно ранговую матрицу X размера $n \times m$ можно представить в виде:

$$X = QR \quad (4)$$

где

- $Q^T Q = E$
- R треугольная матрица размера $m \times m$

При таком подходе из необходимого условия минимума функционала

$$\frac{\partial}{\partial \theta} J(X, h_{\theta}) = 0, \quad (5)$$

следует, что

$$\theta = R^{-1} Q^T y \quad (6)$$

Для реализации данного подхода к задачам, например, по реализации молочных продуктов в животноводстве, где нам больше подходит линейная функция $y(t) = A * t + B$, применение нормального уравнения (3) дает следующие результаты прогноза.

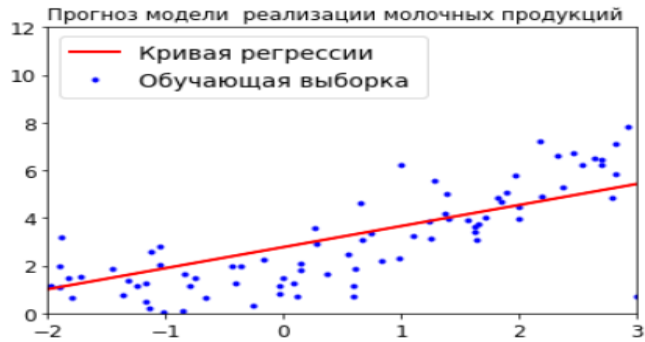


Рисунок 1. Прогноз моделирования молочных продукций.

А для урожайности пшеницы более удобным для нас является полиномиальная функция, например вида $y = 0.5 * X^{**}2 + X + 2$

В данном, случае применение линейной регрессии, например, для данных урожайности пшеницы дает результаты прогноза следующего вида:

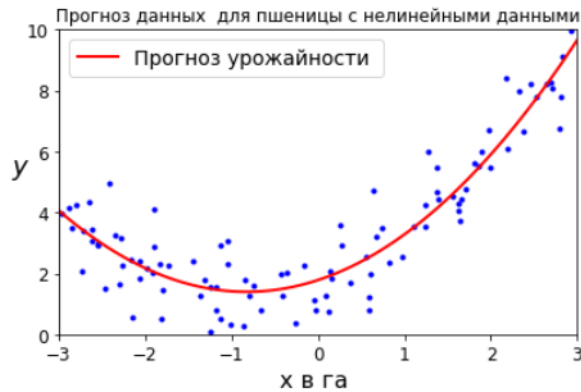


Рисунок 2. Нелинейные данные при прогнозировании моделирования урожайности пшеницы

Далее прогнозная кривая показано ниже

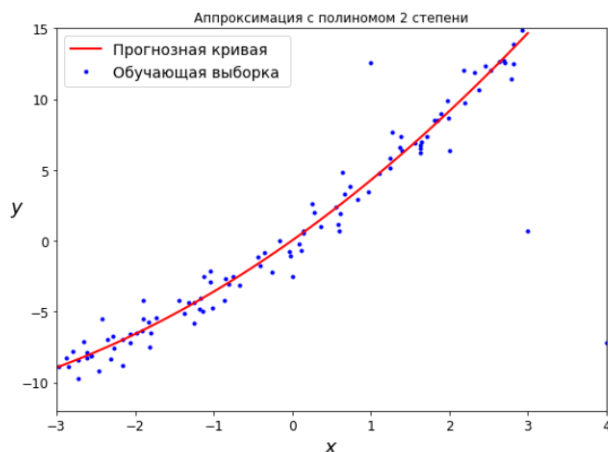


Рисунок 3. Технологии аппроксимации полиномами высокого порядка.

Более сложные случаи встречаются, например, при реализации продукций в животноводстве. Наиболее подходящей в данном случае является логистическая

функция вида $y(t)=A/(1+\exp(B-C*t))$ графическое изображение, которой с шумом имеет вид:

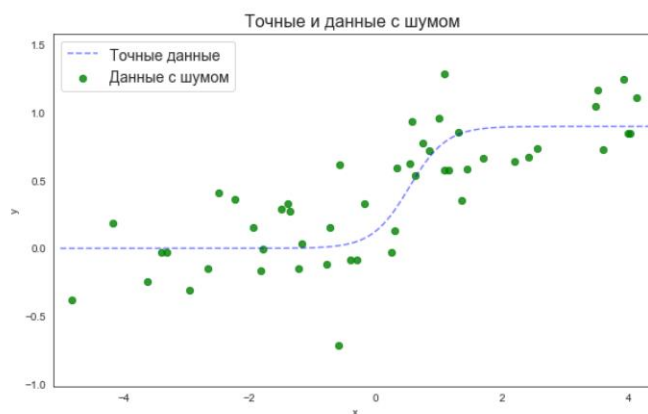


Рисунок 4. Сложные нелинейные данные для моделирования

В этом случае применение технологии линейной регрессии не дает утешительных результатов. Результат применения прогноза полученное регрессией показан на рисунке 5.

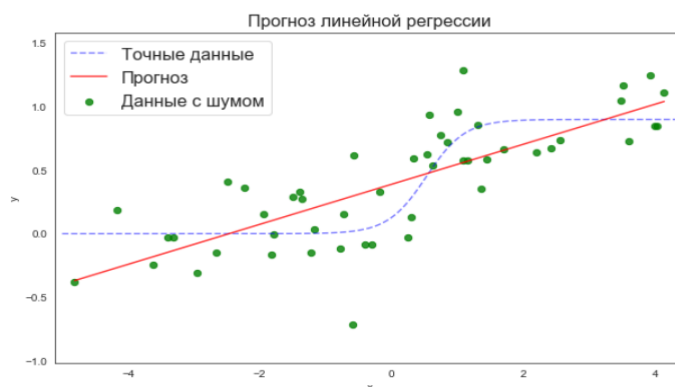


Рисунок 5. Прогнозирование со сложные нелинейные данными.

В таких случаях необходимо применять технологии нелинейного моделирования машинного обучения или применяется метод регуляризации [3] А.Н.Тихонова.

Второй общий подход применение, это численный подход к вычислению параметров модели с применением градиентного метода оптимизации.

Изучается несколько вариантов численного подхода к вычислению минимума функционала.

Метод градиентного спуска для задач оптимизации представляет собой самый общий алгоритм оптимизации, способный находить оптимальные решения широкого диапазона задач[5,6]. Основная идея градиентного спуска заключается в том, чтобы итеративно подстраивать параметры для сведения к минимуму функции издержек, производная которая определяется по формуле

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{2}{m} \sum_{i=1}^m (\theta^T * x^{(i)} - y^{(i)}) x_j^{(i)} \quad (4)$$

Процесс реализации градиентного метода в данном случае определяется по следующей формуле.

$$\theta^{k+1} = \theta^k - \eta \nabla_{\theta} J(\theta^k)$$

Для реализации зерновых и зерново-бобовых культур и построении модели наиболее подходит в данном случае параболическая функция вида $y = 0.25x^2 + x + 2$. Результаты данной реализации показаны на рисунке 6.

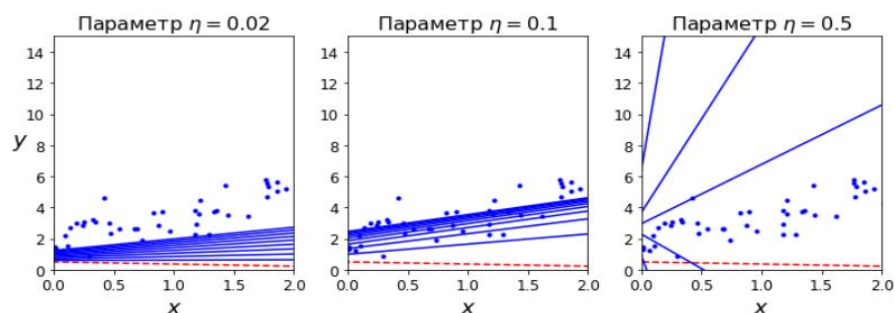


Рисунок 6. Численный подход к прогнозированию задач со сложными нелинейными данными.

Слева скорость обучения модели при $\eta = 0.02$ слишком низкая: в конце концов, алгоритм достигнет решения, но это займет долгое время. На рисунке, посередине при $\eta = 0.1$ скорость обучения выглядит довольно неплохим: алгоритм в этом случае сходится к решению всего за несколько итераций. А вот на рисунке справа при $\eta = 0.5$ скорость обучения очень высокая и наш алгоритм расходится, беспорядочно перескакивая с места на место и фактически с каждым шагом все больше удаляясь от исходного минимума функционала.

Рассмотрена оптимизационная задача применительно к задачам АПК. Изучены варианты самого общего алгоритма -метода градиентного спуска для построения моделей для задач сельского хозяйства, в частности построены модели и прогнозы по реализации сельхозпродукций с использованием регрессионного анализа. Создан пакет программ с использованием интеллектуальных систем

машинного обучения по технологиям Python. Данная методика позволяет изучать и строить модели для широкого круга задач для сферы АПК КР.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. **Х. Бринк, Дж. Ричардс, М. Февееролф.** *Машинное обучение*, -СПб.: Питер, 2018.
2. **А. Мюллер, С. Гвидо.** *Введение в машинное обучение.* -Диалектика, 2019.
3. **А.Н.Тихонов, В.Я.Арсенин.** *Методы решения некорректно поставленных задач.* Наука, Москва, 1978.
4. **Ф.П.Васильев.** *Численные методы решения экстремальных задач,* -Наука, Москва 1978.
5. *Информационная система и математические модели для региональных экономик / С. К. Бийбосунова, Б. Д. Давлятова, У. Т. Керимов, Н. О. Асанбекова // Современные проблемы механики. – 2016. – № 26(4). – С. 75-82.*