

*Журовський Я.О., студент 2
курсу спеціальності 122
«Комп'ютерні науки»
Науковий керівник:
Потапова Н. А., к.е.н., доцент,
доцент кафедри інформаційних
технологій*

АПРОКСИМАЦІЯ ФУНКЦІЙ В DATA SCIENCE

Донецький національний університет імені Василя Стуса, м. Вінниця

В сучасному світі, де кількість доступних даних зростає з кожним днем, виникає потреба в ефективних методах обробки, аналізу та моделювання цих даних. Один із найважливіших етапів у роботі з даними – це апроксимація функцій, тобто знаходження математичних функцій, що найкраще описують набір даних.

У Data Science апроксимація функцій є ключовою задачею, яка забезпечує можливість побудови моделей та прогнозування майбутніх значень на основі наявних даних. Для цього використовуються різні методи апроксимації, такі як лінійна та нелінійна регресія, інтерполяція, сіткові методи, а також нейронні мережі.

Лінійна регресія - це метод апроксимації функцій, який використовується для знаходження лінійної залежності між залежною та незалежними змінними в даних. Лінійна регресія може бути використана для передбачення значення залежної змінної на основі значень незалежних змінних. Залежна змінна може бути числовою, категорійною або бінарною. Лінійна регресія широко використовується в Data Science для передбачення значень цінкових категорій, продажів, прибутку та інших показників.

Нелінійна регресія - це метод апроксимації функцій, який використовується для знаходження нелінійної залежності між залежною та незалежними змінними в даних. Нелінійна регресія може бути використана для передбачення значення залежної змінної на основі значень незалежних змінних, коли залежність не може бути описана лінійною функцією. Нелінійна регресія може використовувати такі функції як експоненційну, логарифмічну, поліноміальну та інші.

Інтерполяція – це метод апроксимації функцій, який використовується для знаходження функції, яка проходить через задані точки. Інтерполяція може бути використана для відтворення даних, які були втрачені або пошкоджені. Інтерполяція може бути лінійною або нелінійною, залежно від типу функції, яку ми спробуємо наблизити.

Лінійна інтерполяція полягає у побудові прямої лінії між двома заданими точками на графіку функції. Цей метод є досить простим, але може

бути менш точним, особливо якщо точки досить далеко одна від одної. Нелінійна інтерполяція використовується тоді, коли ми бажаємо апроксимувати функцію, яка не є лінійною. Найпоширенішим методом нелінійної інтерполяції є многочлен Лагранжа, який використовується для побудови многочлена зі ступеням на один менше, ніж кількість заданих точок. Інший поширений метод нелінійної інтерполяції - це метод найменших квадратів, який використовується для побудови кривої, що найкращим чином відповідає заданим точкам. Інтерполяція може бути корисною, коли ми маємо набір точок і бажаємо побудувати функцію, яка проходить через ці точки. Проте інтерполяція може також призвести до перенавантаження моделі, що може призвести до поганої узгодженості з іншими даними або до недооцінки помилки моделі.

Іншим методом апроксимації функцій є апроксимація найменших квадратів, яка полягає в пошуку функції, яка найбільш точно апроксимує набір точок. Цей метод може бути особливо корисним, коли ми маємо значення, які мають шум або невизначеність, оскільки він може допомогти зменшити вплив цих помилок.

Метод головних компонент (PCA) є ще одним методом апроксимації функцій, який використовується в Data Science. Цей метод дозволяє зменшити розмірність даних, зберігаючи при цьому найбільш значущі ознаки. Основна ідея методу полягає в тому, щоб знайти нові координати для даних, які максимально розподіляють дисперсію даних по цих координатах. Таким чином, ми можемо перетворити дані на новий простір меншої розмірності, де кожна нова координата називається головною компонентою. Головні компоненти є ортогональними між собою та впорядковані за зменшенням відповідних власних значень, які відображають дисперсію даних по кожній головній компоненті.

Метод PCA може бути використаний для зменшення розмірності даних перед виконанням аналізу та моделювання даних. Це зменшує кількість змінних, які несуть більшу частину шуму та дозволяє зберегти тільки найбільш значущі ознаки даних. Це може покращити результати аналізу та моделювання даних, зменшити обчислювальну складність та скоротити час обчислень. Недоліком методу PCA є те, що він може не підходити для даних з високою кореляцією між змінними або для даних з неоднорідними дисперсіями в різних напрямках. Також, втрачається інтерпретованість змінних, оскільки нові головні компоненти є лінійними комбінаціями початкових змінних.

Інший метод - це метод гладких сферичних наближень (Smooth Spherical Approximation), який використовується для апроксимації даних у вигляді сферичної функції. Цей метод дозволяє отримати більш точну апроксимацію даних, особливо для даних, які мають високу ступінь сферичної симетрії.

Інший метод - це метод кускової лінійної апроксимації (Piecewise Linear Approximation), який використовується для апроксимації даних у вигляді лінійних сегментів. Цей метод зазвичай використовується для аналізу та

візуалізації даних, які мають складні форми та треба поділити на більш прості елементи.

Останнім часом все більш популярним стає використання нейронних мереж для апроксимації функцій. Нейронні мережі можуть навчатись на великих обсягах даних та вміти знаходити складні зв'язки між даними. Вони можуть бути використані для передбачення та моделювання даних, а також для зменшення обчислювальної складності.

У загальному, використання апроксимації функцій може допомогти в Data Science при розв'язанні різноманітних завдань, таких як передбачення, класифікація, кластеризація та інші.

Наприклад, апроксимація функцій може бути використана для зменшення розмірності даних. Коли маємо велику кількість ознак, моделі можуть стати надмірно складними, що призводить до погіршення їх точності та збільшення часу обчислень. Застосування методу головних компонент може допомогти зменшити кількість ознак та зберегти більш значущі з них. Це дозволяє зберегти час та ресурси обчислювальної техніки, а також знизити шум у даних, зберігаючи важливі особливості.

Застосування апроксимації функцій є досить поширеним в Data Science та машинному навчанні, оскільки це дозволяє зменшити обчислювальну складність та покращити точність моделей. У мові програмування C# можна використовувати бібліотеку `mathnet.Numerics` для реалізації методів апроксимації функцій. Наприклад, для знаходження поліноміальної апроксимації можна використовувати метод `Polynomial.Fit`.

Усі ці застосування апроксимації функцій показують, що цей метод є потужним інструментом в машинному навчанні та Data Science, який може допомогти в розв'язанні різних завдань та досягненні кращих результатів. Застосування апроксимації функцій дозволяє побудувати апроксимаційну функцію, що наближає вихідну функцію даними з обмеженого набору точок. Апроксимація функцій має широке застосування в багатьох областях, таких як фізика, інженерія, машинний навчання, фінанси, економіка та інші.

Список літератури:

1. Інформаційний портал "MathProfi" - <https://mathprofi.net/aproksimaciya-funkcii.html>
2. Метод найменших квадратів. Апроксимація функцій: навчальний посібник / С.П. Ляшко, В.В. Скобліков, М.А. Ковтун. - К.: КПІ ім. Ігоря Сікорського, 2012. 72 с.
3. Кобзарьов М.В. Методи апроксимації математичних функцій та їх застосування. К.: НТУУ "КПІ", 2008. 139 с.
4. Щоголева О.В., Шевченко А.О. Математичне моделювання в економіці та фінансах. К.: КНЕУ, 2008. - 378 с.