

Линейная регрессия

Введение в машинное обучение

Цель проекта

Цель этого проекта — познакомиться с базовой концепцией машинного обучения.

В ходе данного проекта вам нужно будет создать программу, которая предсказывает цену автомобиля, используя линейную функцию с алгоритмом градиентного спуска.

Мы будем работать с конкретными данными для проекта, но как только вы завершите проект, вы можете использовать алгоритм с любым другим набором данных, если ваша задача подразумевает возможность использования линейной регрессии.

Основные инструкции

Язык программирования – Python. Вы также можете использовать любые библиотеки, которые захотите, но только, если они не делают всю работу за вас. Например, использование `regr.fit(X_train, y_train)` из `sklearn` считается читерством и приведет к неудовлетворительной оценке.

Основная задача

Необходимо реализовать простую линейную регрессию с одной фичей (признаком) – в данном случае, пробегом автомобиля.

Для этого вам нужно создать две программы:

1. Первая программа будет использоваться для прогнозирования цены автомобиля для заданного пробега. При запуске программы она должна запросить у вас пробег, а затем вернуть вам предполагаемую цену (с учетом данного пробега). Программа будет использовать следующую гипотезу для прогнозирования цены:

$$\text{estimatePrice}(\text{mileage}) = \theta_0 + (\theta_1 * \text{mileage})$$

Перед запуском обучения модели θ_0 и θ_1 установлены в 0.

2. Вторая программа будет использоваться для обучения вашей модели. Необходимо прочитать файл с набором данных и построить для них линейную регрессию. После завершения обучения модели, необходимо сохранить переменные θ_0 и θ_1 для использования в первой программе. Предполагается использование следующих формул:

$$\text{tmp}\theta_0 = \text{learnhRate} * \frac{1}{m} \sum_{i=0}^{m-1} (\text{estimatePrice}(\text{mileage}[i]) - \text{price}[i])$$

$$\text{tmp}\theta_1 = \text{learnhRate} * \frac{1}{m} \sum_{i=0}^{m-1} (\text{estimatePrice}(\text{mileage}[i]) - \text{price}[i]) * \text{mileage}[i]$$

Вспомогательные задачи

1. Построить график распределения данных, чтобы лучше понимать, с чем мы имеем дело.
2. Построить график линейной функции, полученной в результате обучения и модели и продемонстрировать, насколько хорошо ваша модель предсказывает стоимость автомобиля.
3. Определить и вывести на экран точность вашей модели.
4. Указать, какой learning rate был использован и почему.

Сдача проекта

Рабочие программы с исходным кодом должны быть размещены в вашей индивидуальной папке на Яндекс Диске (под студенческой учетной записью) Основы Python_Фамилия\Линейная регрессия. В этой же папке необходимо разместить отчет, оформленный средствами LaTeX или Word с описанием вашей работы (необходимо включить формулы, код, выкладки и рассуждения, которыми вы пользовались при реализации проекта).

Необходимо предоставить доступ к указанной папке Основы Python_Фамилия преподавателю.

Время на выполнение – 2 недели. Последний срок размещения проекта на Яндекс Диске **17.11.2024**. Проверка проекта осуществляется очно на занятии.