14.05.2024, 22:33 otchet6 - Colab

Лабораторная работа № 6 по дисциплине "Искусственный интеллект и машинное обучение"

Выполнил: студент 2-го курса Звездин Алексей Сергеевич

Группа: ПИЖ-б-о-22-1

3

Руководитель практики: Березина Виктория Андреевна, ассистент кафедры информационных систем и технологий института цифрового развития

Тема работы: Построение пайплайна одномерной регрессии

Цель работы: Разработка единого пайплайна для решения задачи регрессии

plt.xlabel('Years of Experience')

plt.ylabel('Salary')

plt.show()

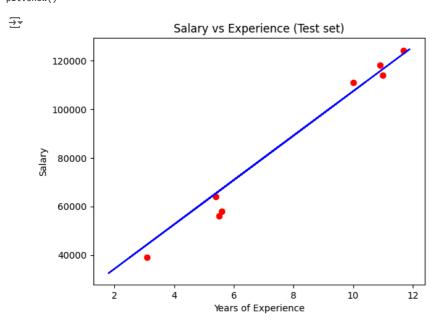
3.5 45000.0 3.8 42000.0

```
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values
print ("Матрица признаков"); print(X[:5])
print ("Зависимая переменная"); print(y[:5])
→ Матрица признаков
     [[1.8]
      [2.5]
      [3.1]
      [3.5]
      [3.8]]
     Зависимая переменная
     [41000. 48000. 39000. 45000. 42000.]
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 1/4, random_state = 0)
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
     ▼ LinearRegression
     LinearRegression()
y_pred = regressor.predict(X_test)
print(y_pred)
→ [ 44404.60502151 122838.85299256 67205.25850147 65381.20622307
      115542.64387898 107334.40862619 116454.67001817 66293.23236227]
plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Salary vs Experience (Training set)')
```

14.05.2024, 22:33 otchet6 - Colab



```
plt.scatter(X_test, y_test, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Salary vs Experience (Test set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```



Контрольные вопросы:

- 1. Почему при реализации линейной модели регрессии нет необходимости выполнять масштабирование признаков
 При реализации линейной модели регрессии нет необходимости выполнять масштабирование признаков, потому что
 коэффициенты регрессии, получаемые в результате обучения модели, будут выражать изменение в зависимой переменной в
 ответ на изменение на единицу независимой переменной. Масштабирование может быть полезно для других моделей, таких
- 2. Почему при реализации модели линейной регрессии в качестве функции потерь используется квадратичное отклонение, а не модуль отклонения?
 - При реализации модели линейной регрессии в качестве функции потерь используется квадратичное отклонение, потому что это позволяет легко вычислить градиент функции потерь и использовать метод градиентного спуска для оптимизации коэффициентов модели. Квадратичное отклонение также обладает хорошими статистическими свойствами, такими как дифференцируемость и выпуклость.
- 3. Что именно реализовано в методе fit(X, y) класса LinearRegression?

как метод ближайших соседей или нейронные сети, но не обязательно для линейной регрессии.

14.05.2024, 22:33 otchet6 - Colab

Метод fit(X, y) класса LinearRegression реализует процесс обучения модели, который заключается в подгонке коэффициентов регрессии таким образом, чтобы минимизировать сумму квадратов остатков между фактическими и предсказанными значениями зависимой переменной.

- 4. Что такое р-значение? Как р-значение используется при оптимизации моделей регрессии?
 - р-значение (p-value) это вероятность получить результаты, более экстремальные, чем фактически наблюдаемые, при условии, что нулевая гипотеза верна. В контексте оптимизации моделей регрессии p-значение используется для оценки статистической значимости коэффициентов регрессии: коэффициенты с низким p-значением считаются статистически значимыми, что означает, что они действительно влияют на зависимую переменную.
- 5. Поясните назначение метода predict класса LinearRegression.
 - Метод predict класса LinearRegression используется для предсказания значений зависимой переменной на основе независимых переменных, для которых модель была обучена с использованием метода fit.
- 6. Поясните назначение метода plot и scatter класса pyplot.
 - Методы plot и scatter класса pyplot из библиотеки matplotlib используются для создания графиков. Метод plot позволяет построить линейный график, а метод scatter используется для построения диаграммы рассеяния, показывающей взаимосвязь между двумя переменными.
- 7. По какой подвыборке необходимо оценивать точность модели машинного обучения: тестовой или тренировочной?
 - Точность модели машинного обучения должна оцениваться на тестовой подвыборке. Тренировочная подвыборка используется для обучения модели, а тестовая подвыборка для оценки ее производительности на новых данных, которые модель ранее не видела.