

Amaliy ish № 7

Mashinali o'qitish modellari

Ishning maqsadi:

Ma'lumotlarni bashorat qilish uchun Scikit-learn kutubxonasidan foydalanib, oddiy mashinali o'qitish modellarini yaratishni o'rganish.

Nazariy qism:

1. Mashinali o'qitish nima?

Machine Learning (ML) - bu sun'iy intellektning bo'limi bo'lib, u kompyuterlarga aniq dasturlashsiz ma'lumotlardan "o'rganish" imkonini beradi.

Mashinali o'qitish vazifasining asosiy turlari:

- Tasniflash: toifalarni bashorat qilish (masalan, omon qolgan / omon qolmagan).
- Regressiya: raqamli qiymatlarni bashorat qilish (masalan, uyning qiymati).
- Klasterlash: o'xshashlik asosida ma'lumotlarni guruhlash.

2. Scikit-learn nima?

Scikit-learn - bu mashinali o'qitish uchun mashhur Python kutubxonasi. U quyidagi vositalarni taqdim etadi:

- Ma'lumotlarni tayyorlash.
- Modellar o'qitish.
- Modellar sifatini baholash.

3. Mashinali o'qitishning asosiy bosqichlari:

- Ma'lumotlarni tayyorlash: ma'lumotlarni tozalash, o'zgartirish va ajratish.
- Modelni o'qitish: O'rgatuvchi ma'lumotlar bo'yicha algoritmnı tanlash va modelni o'qitish.
- Modelni baholash: test ma'lumotlari bo'yicha model sifatini tekshirish.
- Prognozlash: yangi ma'lumotlar haqida bashorat qilish uchun modeldan foydalanish.

Amaliy qism:

1. Kerakli vositalarni o'rnatish:

Python (3.8+ versiyasi) va Pandas va Scikit-learn kutubxonalari o'rnatilganligiga ishonch hosil qiling. Agar ular o'rnatilmagan bo'lsa, buyruqni bajaring:

```
pip install pandas scikit-learn
```

2. Ma'lumotlarni tayyorlash:

- Oldingi ishlarda yuklab olingan Titanic (`train.csv`) ma'lumotlar to'plamidan foydalaning.

3. Ma'lumotlar bilan ishlash:

- Yangi Python faylini yarating (masalan, `amaliy_6.py`) va quyidagi amallarni bajaring:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score

# Ma'lumotlar yuklash
data = pd.read_csv('train.csv')

# Ma'lumotlarni tayyorlash
# Xususiyatlarni tanlash: kabina sinfi, jinsi, yoshi, chipta narxi
X = data[['Pclass', 'Sex', 'Age', 'Fare']]
y = data['Survived'] # Maqsadli o'zgaruvchi: omon qoldi / omon qolmadi

# Kategorik ma'lumotlarni aylantirish (jins -> raqamli qiymatlar)
X['Sex'] = X['Sex'].map({'male': 0, 'female': 1})

# Tushirib qoldirilgan qiymatlarni median bilan to'ldirish
X['Age'].fillna(X['Age'].median(), inplace=True)
X['Fare'].fillna(X['Fare'].median(), inplace=True)

# Ma'lumotlarni o'qitish va sinov namunalariga bo'lish
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
```

```
# Tasodifiy o'rmon modelini o'rgatish
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# Test ma'lumotlari bo'yicha bashorat
y_pred = model.predict(X_test)

# Modelning aniqligini baholash
accuracy = accuracy_score(y_test, y_pred)
print(f"Model aniqligi: {accuracy:.2f}")
```

4. Dasturni ishga tushirish:

Faylni saqlang va uni terminal orqali ishga tushiring:

```
python amaliy_6.py
```

Natijada:

- Modelning aniqligi ko'rinadi (masalan, `Modelning aniqligi: 0,81`).

Qo'shimcha materiallar:

- Rasmiy Scikit-learn hujjatlari: [Scikit-learn Documentation](#).
- Mashinali o'qitish bo'yicha qo'llanma: [Introduction to Machine Learning with Python](#).
- Scikit-learn-dan foydalanishga misollar: [Scikit-learn Tutorials](#).

Amaliy vazifa:

Vazifa: № 6

Titanic (`train.csv`) ma'lumotlar to'plami va Scikit-learn dan foydalanib quyidagi amallarni bajaring:

1. Yo'lovchilarning omon qolishini bashorat qiluvchi logistik regressiya modelini yarating.
2. Testlash ma'lumotlari bo'yicha modelning to'g'riligini baholang.
3. Logistik regressiya modelining aniqligini tasodifiy o'rmon modeli bilan solishtiring.