

№11 Дәріс. Логистикалық регрессия (Logistic Regression)

Дәрістің мақсаты:

- Логистикалық регрессия моделінің мәнін түсіндіру
- Логистикалық функция мен ықтималдық ұғымын түсіндіру
- Модель параметрлерін есептеу принципін көрсету
- Классификация міндеттерінде логистикалық регрессияны қолдануды үйрету
- Python (sklearn) арқылы логистикалық регрессияны жүзеге асыруға негіз дайындау

Кілттік сөздер:

Логистикалық регрессия, бинарлы классификация, сигмоида, ықтималдық, log-loss, odds ratio, decision boundary, maximum likelihood estimation (MLE), confusion matrix.

Дәріс жоспары

1. Кіріспе
 - Классификация деген не?
 - Логистикалық регрессияның қолданылуы
2. Логистикалық регрессияның теориялық негізі
 - Бинарлы классификация
 - Сигмоида функциясы
 - Ықтималдық түсінігі
3. Математикалық модель
 - Логистикалық функция: Sigmoid
 - Odds және log-odds түсініктері
 - Модель формуласы
4. Модель параметрлерін бағалау
 - Maximum Likelihood Estimation (MLE)
 - Cost function (Log loss)
5. Шешім (decision boundary)
6. Модельдің сапасын бағалау
 - Accuracy, Precision, Recall, F1-score
 - Confusion Matrix
7. Python арқылы іске асыру (қисындық түсіндіру)
8. Қорытынды
9. Бақылау сұрақтары
10. Пайдаланылған әдебиеттер

Кіріспе

Қазіргі заман ақпараттық қоғамында деректер көлемі қарқынды өсуде, ал осы деректер негізінде дұрыс шешім қабылдау қажеттілігі күннен-күнге артуда. Денсаулық сақтау жүйесінде ауруды анықтау, банктік секторда клиенттердің төлем қабілеттілігін болжау, цифрлық маркетингте клиенттің

өнімді сатып алу ықтималдығын есептеу, онлайн алаяқтықты анықтау — барлығы деректерді классификациялауға негізделген маңызды міндеттер. Осындай мәселелерді шешуде кеңінен қолданылатын әдістердің бірі — логистикалық регрессия. Бұл модель қарапайымдылығы, түсініктілігі және тиімділігі арқасында машиналық оқытудың базалық әрі әмбебап құралдарының бірі болып табылады. Ол объектінің белгілі бір сыныпқа жату ықтималдығын есептеп, шешім қабылдауға мүмкіндік береді.

Логистикалық регрессия — тек теориялық тұрғыдан ғана емес, сонымен қатар практикалық тұрғыдан да маңызды әдіс. Тіпті күрделі нейрондық желілердің кейбір бөліктерінде логистикалық функция қолданылады. Яғни, бұл тақырып машиналық оқытудың фундаменталдық негіздерінің бірі болып саналады.

Логистикалық регрессия — классификацияның ең танымал және тиімді әдістерінің бірі. Ол объектінің белгілі бір класқа жату ықтималдығын болжайды. Мысалы:

- Клиент несие қайтарады ма? (иә/жоқ)
- Ауру бар ма? (иә/жоқ)
- Хат спам ба? (иә/жоқ)

Логистикалық регрессия қарапайымдылығы, түсіндірілетіндігі және тиімділігі арқасында машиналық оқудың негізін құрайды. Ол нейрондық желілер сияқты күрделі модельдердің де базалық элементі болып саналады.

1. Бинарлы классификация

Объект екі кластың біріне жататынын болжау:

$$y \in \{0, 1\}$$

Бұл жерде:

- 1 = оң нәтиже (positive class)
- 0 = теріс нәтиже (negative class)

2. Сигмоида функциясы (Sigmoid function)

Сызықтық регрессия нәтижесін $(-\infty, +\infty)$ аралығынан $(0,1)$ ықтималдыққа түрлендіреді:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

$$z = b_0 + b_1x_1 + \dots + b_nx_n$$

3. Логистикалық регрессия моделі

$$P(y = 1|x) = \frac{1}{1 + e^{-(b_0 + b_1x)}}$$

4. Odds және Log-Odds

$$Odds = \frac{P}{1 - P}$$

$$Log(Odds) = b_0 + b_1x$$

Параметр бағалау.

Maximum Likelihood Estimation (MLE) арқылы жүзеге асады. Қателік функция — log-loss:

$$Loss = -\frac{1}{N} \sum [y \ln(p) + (1 - y) \ln(1 - p)]$$

Модель сапасын бағалау.

Метрика	Түсінігі
Accuracy	Дұрыс классификация үлесі
Precision	Дұрыс оң нәтижелер үлесі
Recall	Жүйенің оң класты табу қабілеті
F1-score	Precision мен Recall орташа көрсеткіші
Confusion Matrix	Нәтижелер кестесі

Python (sklearn) мысалы

```
python

from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X_train, y_train)
pred = model.predict(X_test)
```

Қорытынды

Логистикалық регрессия — машиналық оқытудың ең негізгі әрі кеңінен қолданылатын әдістерінің бірі. Оның басты құндылығы — модельдің қарапайымдылығы мен түсініктілігі. Бұл әдіс нәтижесін түсіндіру оңай, сондықтан медицина, қаржы, қауіпсіздік, аналитика сияқты салаларда шешім қабылдауды негіздеу үшін жиі қолданылады.

Сонымен қатар логистикалық регрессия — көптеген күрделі алгоритмдердің негізі болып табылатын базалық статистикалық-математикалық механизм. Логистикалық модельдің жұмыс принципін түсінбей, нейрондық желілер, ықтималдықтық классификаторлар, байес әдістерін толық меңгеру қиын.

Сондықтан бүгінгі дәріс машиналық оқытуға бағытталған кәсіби білімнің маңызды кезеңдерінің бірі болып саналады. Бұл тақырыпты меңгеру

сізге болашақта күрделі алгоритмдермен сенімді жұмыс істеуге мүмкіндік береді.

Бақылау сұрақтары

1. Логистикалық регрессия дегеніміз не?
2. Сызықтық регрессиядан айырмашылығы неде?
3. Сигмоида функциясының формуласы қандай?
4. Odds және log-odds деген не?
5. Қандай жағдайларда логистикалық регрессия қолданылады?
6. Log-loss қандай мақсатта қолданылады?
7. Confusion matrix не үшін керек?
8. Precision мен Recall айырмашылығы неде?

Пайдаланылған әдебиеттер

1. James G., Witten D., Hastie T., Tibshirani R. *An Introduction to Statistical Learning*. Springer.
2. Bishop C. *Pattern Recognition and Machine Learning*. Springer, 2006.
3. Andrew Ng. *Machine Learning Course Notes*, Stanford University.
4. Géron A. *Hands-On Machine Learning with Scikit-Learn & TensorFlow*. O'Reilly, 2019.
5. Scikit-learn documentation: *Logistic Regression Guide*.