

От регрессии к классификации

Классификация — вид задач машинного обучения по прогнозу значений категориального целевого признака по значениям других. Пример задачи классификации — определение цвета фона на изображении.

Базовый метод решения задач классификаций — **логистическая регрессия**.

```
from sklearn.linear_model import LogisticRegression

# приводим значения всех признаков к единой шкале
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# создаём объект для работы с моделью
clf = LogisticRegression()

# настраиваем модель на данные
clf = clf.fit(X_train_scaled, y_train)

# получаем прогнозные значения
y_pred = clf.predict(X_test_scaled)

# получаем оценки вероятностей для класса
y_proba = clf.predict_proba(X_test_scaled)
```

Порог классификации

Порог классификации — вероятность, достаточная для отнесения объектов к классу 1. По умолчанию равно 0.5. Все наблюдения с оценкой вероятности ниже порога будут отнесены к классу 0, остальные — к классу 1.

```
from sklearn.linear_model import LogisticRegression

# создание и обучение модели
clf = LogisticRegression()
clf = clf.fit(X_train_scaled, y_train)

# предсказание вероятностей
y_proba = clf.predict_proba(X_test_scaled)[: , 1]
```