Отбор признаков и интерпретация моделей

Отбор признаков (англ. feature selection) — это оценка важности каждого входного признака алгоритмами машинного обучения. За отбором признаков следует удаление избыточных.

- Отбор признаков эффективно сокращает сложность модели, повышая её предсказательные свойства.
- Отбор признаков уменьшает расход ресурсов на обучение, даже если оставляет качество на прежнем уровне.

Интерпретация моделей — объяснение того, как модель принимает решения и делает предсказания.

Бизнесу важно понимать, как «рассуждает» модель, чтобы грамотно рассчитывать последствия от её внедрения. Любая ошибка чревата убытками и стратегическими просчётами — используйте все доступные средства, чтобы их предотвратить.

Методы отбора признаков

- Метод фильтрации проводит анализ дисперсий признаков и указывает на те, у которых низкая вариативность.
- L1-регуляризация помогает убрать признаки, коэффициенты которых обнулила регуляризация.
- SelectKBest составляет рейтинг признаков, которые вносят наибольший вклад в модель.
- Введение шума добавляет «шумный» признак в модель и сможет отобрать среди исходных признаков те, влияние которых близко к шумному.

Методы интерпретации модели

SHAP — метод оценки важности признаков, который вычисляет, как отдельный признак помог присвоить модели итоговое значение целевого.

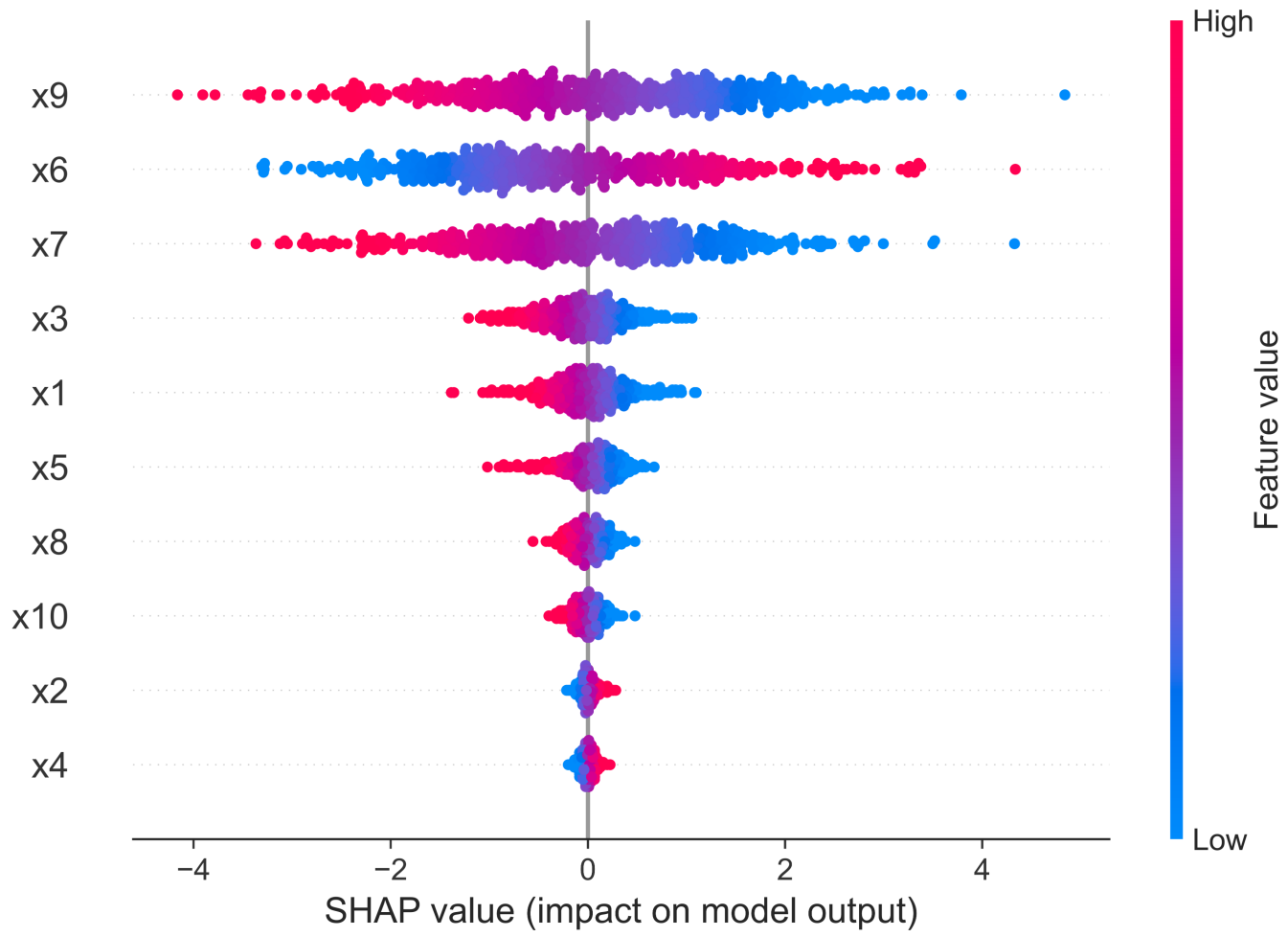
Значения SHAP можно вывести на разных графиках:

- beeswarm показывает общую важность признаков модели по значениям Шепли, присвоенным всем наблюдениям в выборке.

```
import shap

explainer = shap.LinearExplainer(model, X)
shap_values = explainer(X)

shap.plots.beeswarm(shap_values)
```



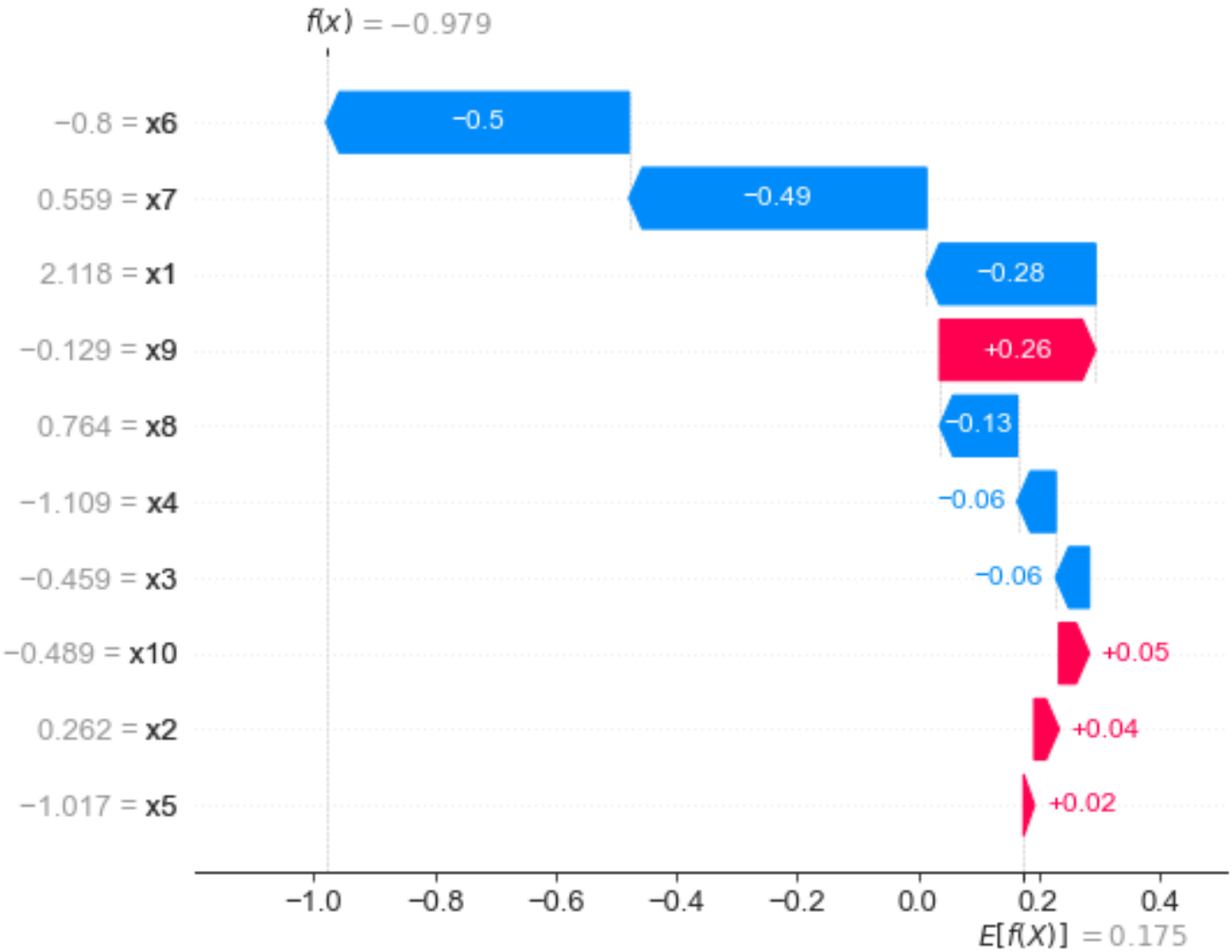
Отбор признаков и интерпретация моделей

- waterfall визуализирует индивидуальные SHAP-значения отдельных наблюдений в датасете.

```
import shap

explainer = shap.LinearExplainer(model, X)
shap_values = explainer(X)

shap.plots.waterfall(shap_values[5])  # график для пятого объекта
```



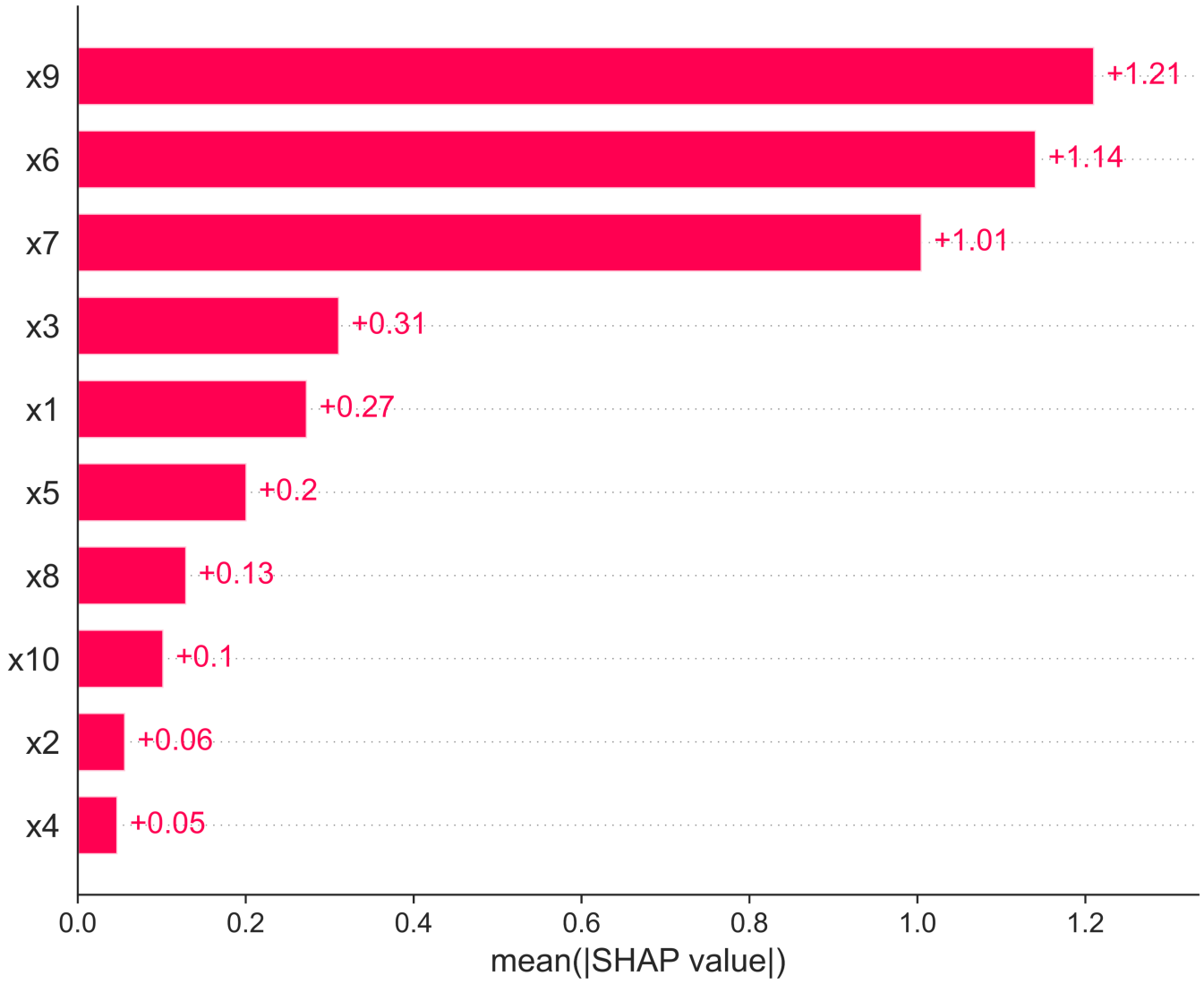
Отбор признаков и интерпретация моделей

- bar отражает агрегированный вклад признаков в прогнозы модели. Он показывает средние SHAP-значения признака по всем наблюдениям. Для расчёта средних берут значения Шепли по модулю, чтобы положительные и отрицательные значения не сводили друг друга к нулю.

```
import shap

explainer = shap.LinearExplainer(model, X)
shap_values = explainer(X)

shap.plots.bar(shap_values)
```



Отбор признаков и интерпретация моделей

Permutation importance — метод оценки важности признака с помощью перемешивания его значений. Если после перемешивания коэффициенты признака в модели стали ниже — значит, он важен.

```
from sklearn.inspection import permutation_importance

permutation = permutation_importance(
    model,  # обученная модель
    X_test,  # тестовые данные
    y_test,  # тестовый целевой признак
    scoring=f1  # метрика для оценивания
)
```