

BIOF3 FIGCODE · SCI 绘图代码集

SHAP 特征重要性 (SHAP Summary Plot)

导出日期: 2026年6月27日



扫码关注微信公众号【生信F3】
获取文章完整内容，分享生物信息学
最新知识。

分类: 分布可视化 | 依赖包: xgboost、shap、scikit-learn、matplotlib

解决的生物学问题

我的机器学习模型是被哪些特征驱动的? 高表达基因 X 推高还是拉低预测概率?

应用场景

- ML 模型可解释性
- 特征重要性排序
- 论文回答审稿人"模型为什么这样预测"
- 黑盒模型变白盒

输入数据格式

sklearn / xgboost 训练好的模型 + 特征矩阵

关键参数

- max_display (展示 top N 特征)
- plot_type = "dot" / "bar" / "violin"
- feature_names

快速开始

1. 直接在线运行

打开 [FigCode 在线绘图](#)，点击本工具卡片的"在线绘图"按钮即可用内置示例数据出图，零环境配置。

2. 本地复现

下载脚本和示例数据，本地 RStudio 运行：

```
# 下载
curl -O https://<your-site>/figcode/scripts/shap-summary.R
curl -O https://<your-site>/figcode/data/shap-summary.csv
```

3. 安装依赖

```
# CRAN 包
install.packages(c("xgboost", "shap", "scikit-learn", "matplotlib"))

# Bioconductor 包 (如需)
# BiocManager::install(c())
```



SHENGXIN F3



SHENGXIN F3

完整代码

```
import numpy as np
import xgboost as xgb
import shap
import matplotlib.pyplot as plt
from sklearn.datasets import load_breast_cancer

# --- Demo data ---
data = load_breast_cancer()
X, y = data.data, data.target
feature_names = data.feature_names

# --- Train XGBoost ---
model = xgb.XGBClassifier(n_estimators=200, max_depth=4,
                           use_label_encoder=False, eval_metric="logloss")
model.fit(X, y)

# --- SHAP values ---
explainer = shap.Explainer(model, X)
shap_values = explainer(X)

# --- Summary plot (beeswarm) ---
plt.figure(figsize=(9, 7))
shap.summary_plot(
    shap_values, X,
    feature_names=feature_names,
    max_display=15,
```

```
    plot_type="dot",
    show=False
)
plt.title("SHAP feature importance – Breast Cancer")
plt.tight_layout()
plt.savefig("plot_001.png", dpi=150, bbox_inches="tight")
print(f"Trained on {X.shape[0]} samples × {X.shape[1]} features")
```

替换为自己的数据

脚本中以 `# --- Demo data ---` 标注的段落是示例数据生成代码。替换为自己的数据时，**保持列名一致**即可：

- 输入格式：sklearn / xgboost 训练好的模型 + 特征矩阵
- 使用 `read.csv()` / `readRDS()` 读取本地文件

延伸阅读

- 相关教程：[SHAP 特征重要性 完整流程](#)
- 出现 bug？欢迎在 [FigCode 页面](#) 点击对应卡片，在评论区留言。