

Exercício

Classificação e Interpretação de Variáveis com Regressão Logística e SHAP

Objetivo

Implementar um modelo de classificação usando Regressão Logística e interpretar as contribuições das variáveis para o modelo usando SHAP values. O objetivo é que você desenvolva habilidades em construção de modelos de classificação, avaliação de desempenho e interpretação de variáveis.

Descrição do Problema

Vamos utilizar um conjunto de dados real, disponível na biblioteca `sklearn`, chamado **Breast Cancer Wisconsin**. Esse dataset é muito usado em problemas de classificação e contém variáveis relacionadas a características de tumores de mama, com a meta de classificar os tumores como "malignos" ou "benignos". Leia na documentação da biblioteca sobre este conjunto de dados para conhecê-lo melhor. Sua tarefa é construir um modelo de Regressão Logística para classificar os tumores e, em seguida, interpretar a importância das variáveis no modelo usando SHAP values.

Etapas do Exercício

1. Importação das bibliotecas necessárias:

```
# Importação das bibliotecas necessárias
import pandas as pd
import numpy as np
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, roc_auc_score
import shap
import matplotlib.pyplot as plt
import seaborn as sns
```

2. Exploração e Preparação dos Dados:

- Carregue o conjunto de dados .
- Realize uma análise exploratória dos dados para entender a estrutura e as distribuições das variáveis.
- Realize o pré-processamento dos dados conforme necessário, como tratamento de valores ausentes e normalização das variáveis.

```
# Exemplo de como carregar o conjunto de dados Breast Cancer Wisconsin da
# biblioteca sklearn
dados = load_breast_cancer()
X = pd.DataFrame(dados.data, columns=dados.feature_names)
y = pd.Series(dados.target, name="diagnosis") # 0 = Maligno, 1 = Benigno
```

3. Divisão dos Dados:

- Divida o conjunto de dados em treinamento (80%) e teste (20%).

```
# Divisão dos dados em treino e teste
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
```

4. Construção do Modelo de Regressão Logística:

- Implemente um modelo de Regressão Logística para prever a probabilidade de um cliente adquirir o produto.
- Treine o modelo com os dados de treinamento.
- Avalie o desempenho do modelo no conjunto de teste, utilizando métricas como AUC-ROC e acurácia.

5. Interpretação do Modelo com SHAP values:

- Utilize a biblioteca SHAP para calcular os SHAP values do modelo.
- Interprete quais variáveis têm o maior impacto na probabilidade de um cliente realizar uma compra.
- Explore, pelo menos, três exemplos de clientes para visualizar como as variáveis individuais influenciam a decisão do modelo.

```
# Inicializando o Explainer SHAP para o modelo de Regressão Logística
explainer = shap.Explainer(modelo, X_train)
shap_values = explainer(X_test)

# Plot SHAP summary para entender a importância geral das variáveis
shap.summary_plot(shap_values, X_test, plot_type="bar")

# Análise detalhada para três exemplos de pacientes
for i in range(3):
    print(f"\nInterpretação para o exemplo {i+1}:")
    shap.initjs()
    shap.waterfall_plot(shap_values[i])
```

6. Análise e Discussão dos Resultados:

- Com base nos SHAP values, discuta quais variáveis foram mais influentes no modelo e se os resultados são coerentes com a expectativa inicial.
- Escreva uma análise de como as variáveis mais impactantes poderiam orientar futuras campanhas de marketing.
- Entregue um PDF com tudo o que foi solicitado acima e sem os códigos.

Informações adicionais

Sobre o gráfico Waterfall

O gráfico waterfall do SHAP (SHapley Additive exPlanations) é uma excelente ferramenta para interpretar as previsões de um modelo em relação a uma observação individual. Ele mostra como cada variável (ou feature) contribui positiva ou negativamente para a previsão final do modelo em relação ao valor médio.

Estrutura do Gráfico Waterfall SHAP

1. Base Value (Valor Base)

- O gráfico começa com o "base value" (valor base), que é a previsão média do modelo para o conjunto de dados. Esse valor representa o ponto de partida antes de considerar qualquer uma das variáveis específicas do exemplo.
- No contexto de um problema de classificação, como neste exemplo, o "base value" é a probabilidade média de uma das classes (exemplo: benigno ou maligno).

2. Contribuições das Variáveis

- Cada linha representa uma variável e indica o quanto essa variável específica adiciona ou subtrai do valor base para chegar ao valor final da previsão.
- Se a variável tiver uma contribuição positiva, ela "empurra" o valor em direção à direita no gráfico (em direção à classe-alvo com probabilidade 1).
- Se a variável tiver uma contribuição negativa, ela "empurra" o valor em direção à esquerda (em direção à classe oposta ou com probabilidade 0).

3. Magnitude e Sinal dos SHAP Values

- A largura de cada barra indica a magnitude do impacto de uma variável. Quanto maior a barra, maior é o impacto daquela variável na previsão final.
- O sinal (direção da barra) indica se a variável está contribuindo para aumentar ou diminuir a previsão. Por exemplo:
 - Uma variável que empurra a previsão para cima indica que aumenta a probabilidade de uma classe (por exemplo, "maligno" no caso de tumores).
 - Uma variável que empurra para baixo indica que diminui a probabilidade de uma classe (por exemplo, "benigno" no caso de tumores).

4. Expected Value (Valor Esperado)

- No final do gráfico, é mostrado o "expected value" ou "valor esperado", que é a previsão final do modelo para a observação em questão, após considerar todas as contribuições das variáveis. No gráfico waterfall de SHAP, o valor $E[f(\mathbf{x})]$, também conhecido como o **base value** ou **valor esperado**, representa a **previsão média do modelo** para o conjunto de dados. Esse valor é o ponto de partida para calcular a previsão de uma observação específica, considerando as contribuições de cada variável. O valor $E[f(\mathbf{x})]$, ou valor base, é essencial porque ele nos fornece uma referência de **onde o modelo começa a prever** antes de levar em conta as características individuais de uma amostra específica. Em termos simples:
 - $E[f(\mathbf{x})]$ é a média das previsões do modelo para o conjunto de dados (caso seja um problema de classificação, pode ser a probabilidade média da classe-alvo).
 - Para uma observação individual, o modelo começa a partir desse valor base e depois ajusta para cima ou para baixo, dependendo das contribuições (SHAP values) das variáveis da amostra.
- Esse valor esperado representa a probabilidade calculada para a classe-alvo da observação analisada.

- No gráfico waterfall de SHAP, o valor **$f(x)$** , exibido no topo, representa a **previsão final do modelo** para a observação específica em análise (ou seja, o "ponto em questão").

No caso do **Breast Cancer Wisconsin dataset** no `sklearn`, o modelo de **Regressão Logística** normalmente classifica as observações em duas classes: **maligno (0)** e **benigno (1)**.

Para um gráfico SHAP com valores de $f(x)$ variando entre negativos e positivos:

- **Valores negativos de $f(x)$** normalmente indicam que o modelo está mais inclinado a prever a classe **maligno (0)**.
- **Valores positivos de $f(x)$** indicam uma inclinação para prever a classe **benigno (1)**.

Interpretação dos Valores de $f(x)$

Em termos gerais:

- Quando o valor de $f(x)$ está próximo de zero, a previsão do modelo está indecisa entre benigno e maligno.
- Valores significativamente negativos indicam uma alta confiança na previsão "maligno".
- Valores positivos indicam uma alta confiança na previsão "benigno".