

Relatório LAMIA - Módulo 23

O vídeo se inicia com definições de classificação e regressão, onde a classificação serve para categorizar objetos, como classificar imagens em um dataset tipo CIFAR-100. Enquanto os modelos de regressão servem para realizar previsões de números utilizando dados passados, por exemplo: valor de uma ação daqui 4 meses com base na variação dos meses passados.

É abordado novamente o conceito de matriz de confusão, feito para avaliar falsos positivos e falsos negativos, porém desta vez foi aprofundado, mostrando a diferença entre precisão e acurácia.

Acurácia: soma de todos os acertos dividida pela quantidade de tentativas

$$\frac{\text{Acertos}}{\text{Total}} = \frac{VN + VP}{VN + FN + VP + FP}$$

precisão: positivos dividido pela soma de falsos positivos e positivos verdadeiros, útil para casos com muitos falsos positivos, onde a acurácia se prova ineficaz para avaliar o modelo

$$\frac{VP}{VP + FP}$$

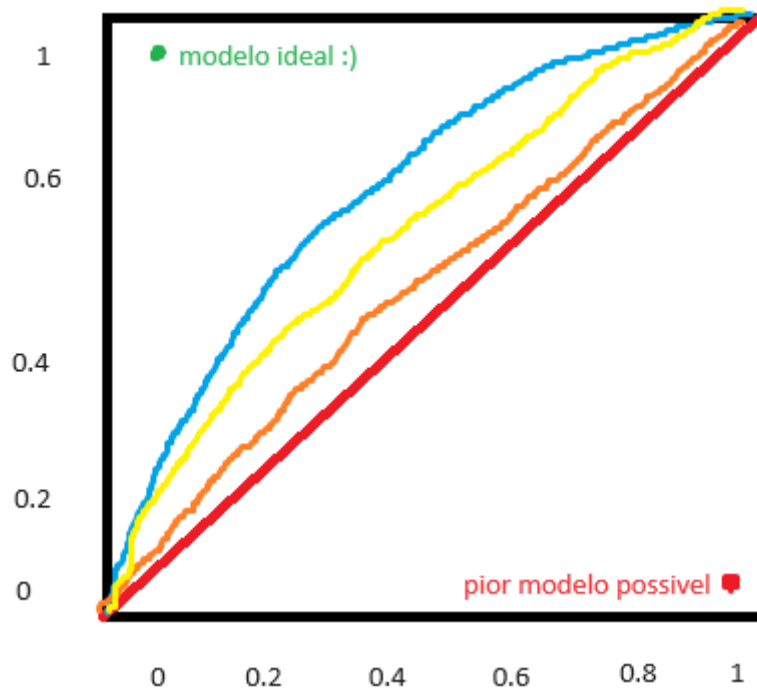
Recall: positivos dividido pela soma de positivos verdadeiros e falsos negativos, útil para minimizar falsos negativos

$$\frac{VP}{VP + FN}$$

Especificidade: Negativos verdadeiros dividido pela soma de negativos verdadeiros e falsos positivos, avalia se o modelo é capaz de rejeitar instâncias negativas

$$\frac{VN}{VN + FP}$$

Em seguida o instrutor do curso menciona a curva ROC, um gráfico com o intuito de relacionar especificidade e sensibilidade, o melhor resultado da curva reside nos valores 0 para especificidade e 1 para sensibilidade



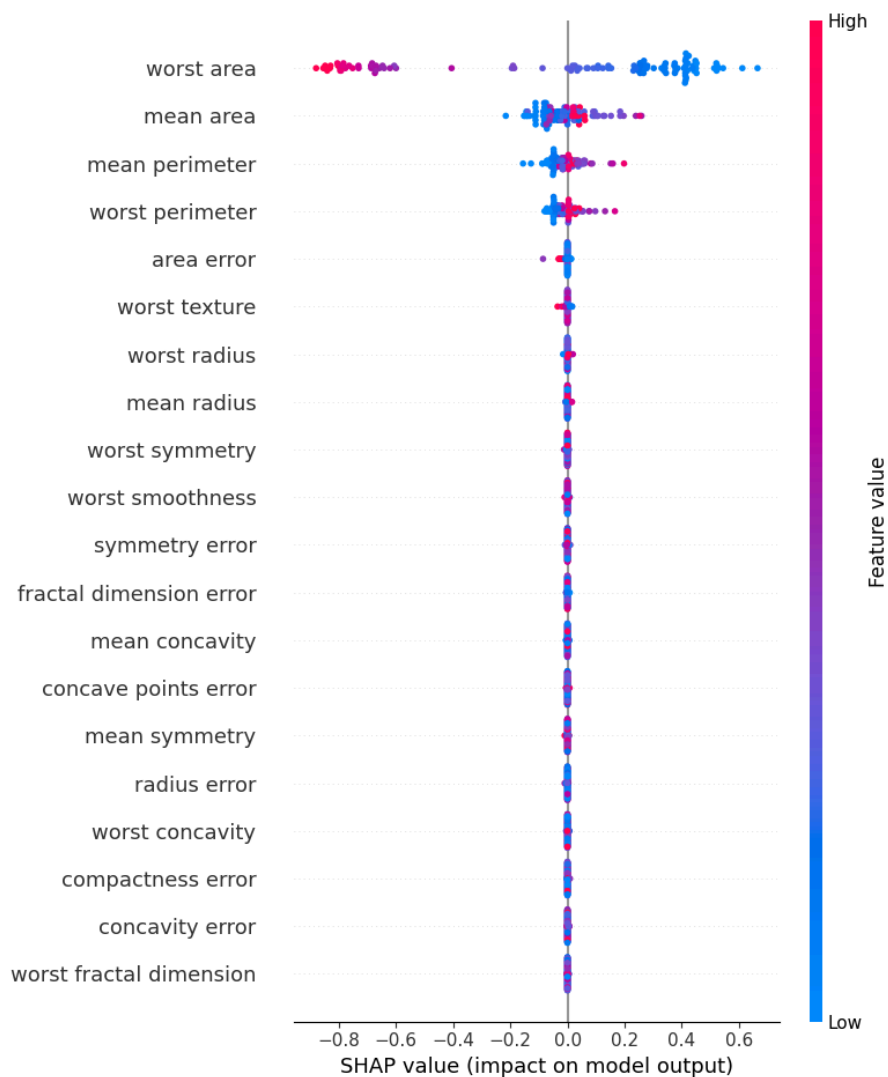
Representação gráfica de uma curva ROC feito em software profissional (Paint)

A representação acima retrata uma curva ROC, onde a curva azul é o melhor modelo dentre as curvas, e o vermelho o pior dentre os apresentados.

SHAP

SHAP é uma ferramenta utilizada para avaliar a contribuição de cada característica no treinamento e teste do modelo, possui diversas configurações, que mudam o cálculo feito para se obter os valores da visualização, alguns exemplos são o kernel explainer, útil para todos os tipos de modelo, e o deep explainer, útil para redes neurais profundas.

É possível obter outros tipos de visualização além da contribuição de cada característica, que é obtida pelo `summary_plot`, é possível visualizar a contribuição calculada pelo SHAP comparada com a contribuição real do modelo para avaliar o impacto da característica.



Exemplo de um SHAP de um modelo Linear SVC de 100 epochs no dataset breast cancer

Para fins de experimentação foi aplicado image plot do SHAP para analisar um modelo para classificação do fashion mnist, avaliando os pixels mais importantes da imagem.

```
shap.image_plot([shap_values], X_test[:10]) # Os pixels vermelhos são os mais influentes, e os azuis influentes de forma
# na maioria dos casos apenas os contornos são analisados
```

