Aprendizagem de Máquina

Seleção de Atributos e Interpretabilidade

Prof. Luiz Eduardo S. Oliveira

Universidade Federal do Paraná
Departamento de Informática
www.inf.ufpr.br/lesoliveira

Introdução

- Um dos principais aspectos na construção de um bom classificador é a utilização de características discriminantes.
- Não é difícil encontrar situações nas quais centenas de características são utilizadas para alimentar um classificador.
- A adição de uma nova característica não significa necessariamente um bom classificador.
  - Depois de um certo ponto, adicionar novas características pode piorar o desempenho do classificador.
- Outro aspecto importante consiste em entender a contribuição de cada característica

Introdução

- Aspectos diretamente relacionados com a escolha das características:
  - Desempenho
  - Tempo de aprendizagem
  - Tamanho da base de dados
- Seleção de características
  - Tarefa de identificar e selecionar um subconjunto de características relevantes para um determinado problema, a partir de um conjunto inicial
    - ⋆ Características relevantes, correlacionadas, ou mesmo irrelevantes.

Introdução

- Não é um problema trivial
  - Em problemas reais, características discriminantes não são conhecidas a priori.
  - Características raramente são totalmente independentes.
  - Duas características irrelevantes, quando unidas pode formar uma nova característica relevante e com bom pode de discriminação.

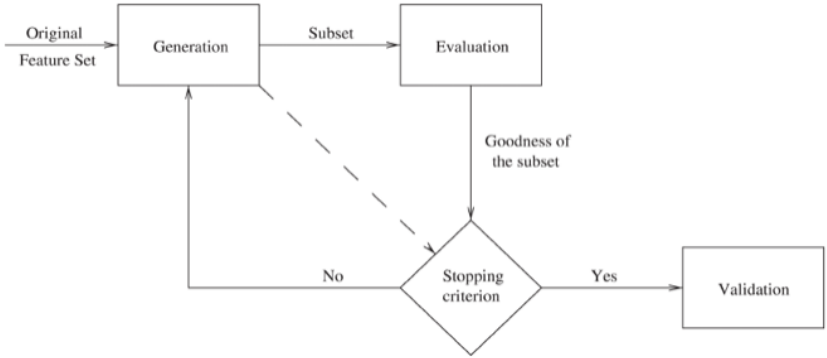
Objetivos

- Encontrar um subconjunto que pode ser:
  - Ideal
    - ⋆ O menor subconjunto necessário e suficiente para resolver um dado problema
  - Clássico
    - ⋆ Selecionar um subconjunto de M características a partir de N características, na qual $M < N$, de maneira a minimizar uma dada função objetivo.
  - Melhor desempenho
    - ⋆ Buscar um subconjunto que melhore o desempenho de um dado classificador.

Visão Geral

- Um método de seleção de características deve utilizar um método de busca para encontrar um subconjunto M a partir de N características
  - Espaço de busca é 2^N
- Para cada solução encontrada nessa busca, uma avaliação se faz necessária.
- Critério de parada
- Validação

Visão Geral


Original
Feature Set
Generation
Subset
Evaluation
Goodness of
the subset
Stopping
criterion
No
Yes
Validation

Gerando subconjuntos candidatos

- Existem diferentes abordagens que podem ser usadas para gerar os subconjuntos
  - Exaustiva
    - Explora todas as possíveis combinações do espaço de busca (2^N)
    - Garante que o subconjunto ótimo será encontrado.
    - Custo computacional elevado e inviável quando o espaço de busca é grande.
  - Heurísticas
    - Forward Selection
    - Backward Elimination
  - Computação evolutiva
    - Algoritmos Genéticos
    - Particle Swarm Optimization

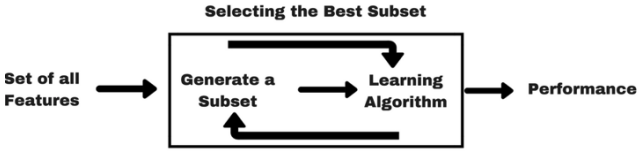
Funções de Avaliação

- Para julgar se um dado subconjunto é ótimo, temos que avaliar o mesmo.
- As funções de avaliação podem ser divididas em:
  - Filter: Independentes do algoritmo de aprendizagem.
  - Wrapper: Dependente do algoritmo de aprendizagem.
  - Embedded: Usa algoritmos capazes de avaliar o poder de discriminação de características.

Métodos Filter

- Geralmente usados como uma passo de pré-processamento
- Independente de qualquer algoritmos de aprendizagem
- Importância das características são medidas através de diversos testes estatísticos, por exemplo, características com pouca variância.
- Sklearn implementa alguns desses métodos na classe **feature_selection**

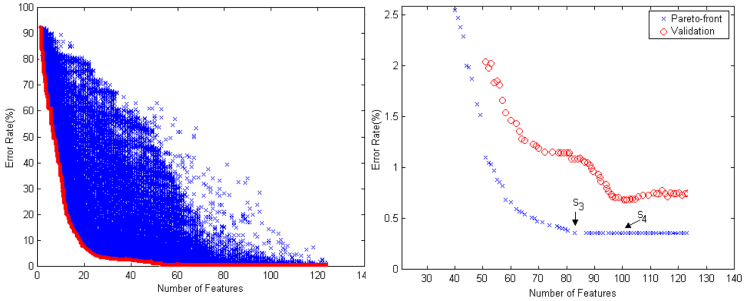
Métodos Wrapper



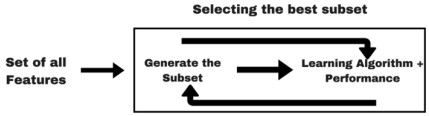
- Em geral produz melhores resultados do que métodos filter
- Custo computacional mais alto
- É importante ter em mente que o processo de seleção de características deve ser visto como um processo de aprendizagem
- Sendo assim, é importante utilizar uma base de validação para evitar over-fitting.
- Quando possível utilize uma base diferente de todas para calcular a função de avaliação

Métodos Wrapper

- Devido ao poder de explorar grandes espaços de busca, algoritmos genéticos tem sido largamente utilizados em problemas de seleção de características
- Um objetivo (desempenho ou um índice qualquer)
- Múltiplos objetivos (quantidade de características, desempenho, etc..)



Métodos Embedded



- Combina as qualidades as abordagens filter e wrapper.
- Utiliza algoritmos de aprendizagem que tem a capacidade de avaliar características, como por exemplo, árvores de decisão.

```
>>> from sklearn.ensemble import ExtraTreesClassifier
>>> from sklearn.datasets import load_iris
>>> from sklearn.feature_selection import SelectFromModel
>>> iris = load_iris()
>>> X, y = iris.data, iris.target
>>> X.shape
(150, 4)
>>> clf = ExtraTreesClassifier(n_estimators=50)
>>> clf = clf.fit(X, y)
>>> clf.feature_importances_
array([ 0.04...,  0.05...,  0.4...,  0.4...])
>>> model = SelectFromModel(clf, prefit=True)
>>> X_new = model.transform(X)
>>> X_new.shape
(150, 2)
```

Exemplo da classe SelectFromModel implementado no Sklearn

Análise de Componentes Principais (PCA)

- Uma ferramenta que pode ser utilizada para **redução de dimensionalidade**
- A idéia é aplicar PCA na base de aprendizagem e encontrar os principais autovetores da base.
- Abordagem filter, visto que o algoritmo de aprendizagem não é utilizado.
- Note que após o PCA, os dados se encontram em um novo espaço de representação.
- Apesar de uma possível redução, todas as características devem continuar sendo extraídas.
- O custo da extração de características não é alterado (somente o custo do algoritmo de aprendizagem)

Como interpretar as características?

- Capacidade de entender e explicar como um modelo toma suas decisões ou faz previsões
- É um aspecto crucial para garantir a confiança, a transparência e a responsabilidade nos sistemas de aprendizado de máquina
  - Especialmente em aplicações críticas como medicina, finanças e justiça.
- Métodos Intrínsecos
  - Árvores de decisão: Cada decisão é baseada em uma série de regras simples e claras.
  - Regressão Linear/Logística: As relações entre as variáveis são representadas por coeficientes que indicam a importância de cada característica.
- Métodos Pós-Hoc
  - SHAP (SHapley Additive exPlanations): Usa valores de Shapley para explicar as contribuições de cada característica em uma predição.
  - LIME (Local Interpretable Model-agnostic Explanations): Cria explicações locais aproximando o comportamento do modelo complexo com um modelo interpretable simples.

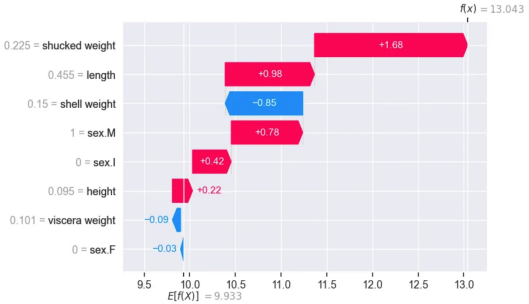
SHAP Values

- Os valores de Shapley, originados da teoria dos jogos, são usados para atribuir importância às contribuições de cada característica (ou variável) em um modelo de aprendizado de máquina.
- Eles fornecem uma forma justa de atribuir crédito a cada característica com base em seu impacto no resultado do modelo.
- Para calcular os valores de Shapley de uma característica específica, são consideradas todas as possíveis combinações das características. Isso é muito caro computacionalmente.
- Uma aproximação que torna o cálculo mais viável é o SHAP (SHapley Additive exPlanations)

Representações Gráficas

Waterfall

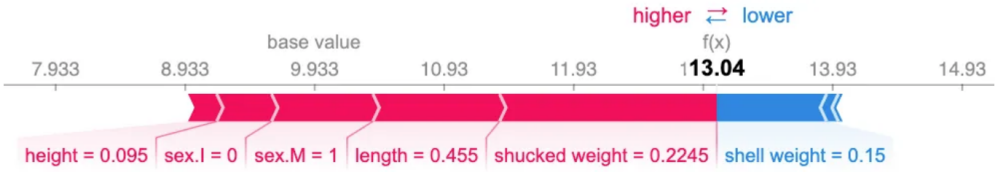
- Esse gráfico mostra os valores SHAP para uma dada observação da base de dados.
- $E[f(x)] = 9.933$ é o valor médio das predições do modelo para toda base.
- $f(x) = 13.043$ é a predição para um dado exemplo
- Os valores SHAP mostram quanto cada característica contribui
  - 13.043 - 9.933 = 3.11
  - 1.68+0.98-0.85+0.78+0.42+0.22-0.09-0.03 = 3.11



Representações Gráficas

Force Plot

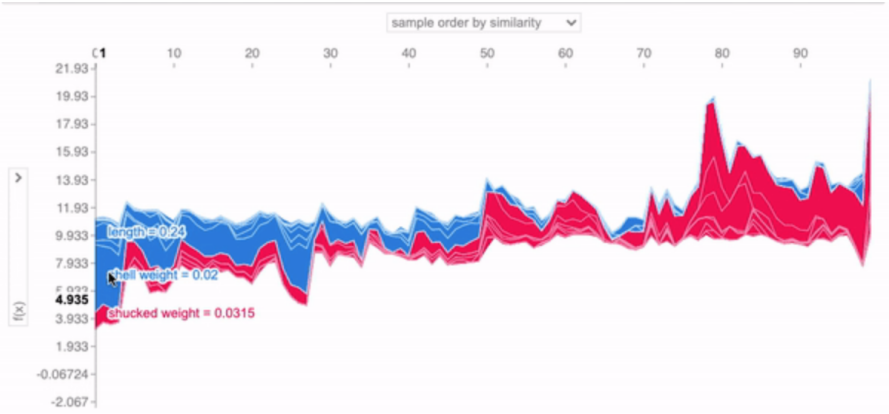
- Esse gráfico pode ser visto como o gráfico anterior condensado.
- Começando no valor base (9.933), é possível visualizar o quanto cada característica contribui para a valor final de predição (13.04)



Representações Gráficas

Stacked Force Plot

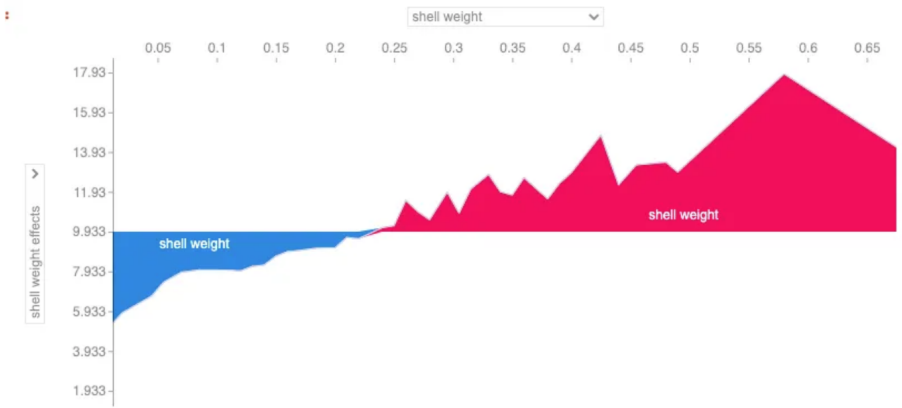
- Os dois gráficos anteriores são usados para analizar uma predição.
- O Stacked Force Plot combina vários Force Plot
  - Gráfico interativo (é possível escolher a variável de interesse)



Representações Gráficas

Stacked Force Plot

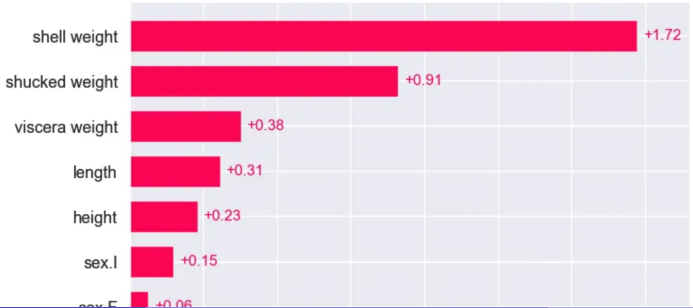
- No exemplo abaixo a variável "shell weight" foi escolhida
- Nesse caso, pode-se observar que valores maiores dessa variável aumentam os Shap Values



Representações Gráficas

Mean Shap

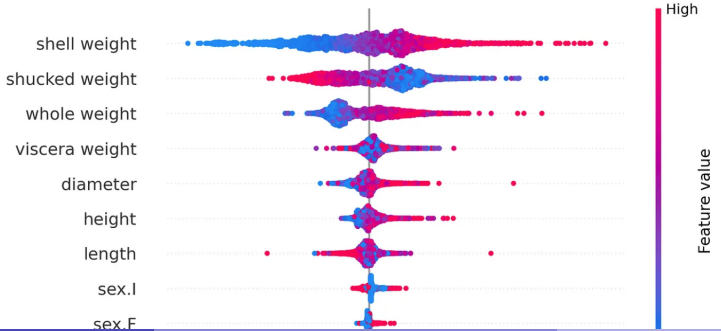
- Mostra as características mais importantes.
- O valor médio de todas as observações é utilizado.
- Características com maiores contribuições (positiva/negativa) apresentam maiores valores médios.
- Muito útil para aqueles modelos que não possuem o atributo de "feature_importance", por exemplo, SVM.



Representações Gráficas

BeeSwarm

- Mostra todos os SHAP Values
- No eixo y, os valores são agrupados por característica. Para cada grupo, a cor doss pontos é determinado pelo valor da característica (vermelho $\rightarrow$ maior)
- Impacto na predição
  - Por exemplo, valores maiores para "shell weight" tem um valor maior na predição. O contrário é observado para "shucked weight"



Local Interpretable Model-agnostic Explanations (LIME)

- Explica uma instância especifica dos dados.
- Para isso, cria um conjunto de dados artificiais ao perturbar a instância original.
  - Isso é feito gerando variações da entrada e obtendo as previsões do modelo para essas variações.
- Com base nas previsões do modelo para os dados perturbados, o LIME ajusta um modelo interpretable e simples, como uma regressão linear, que se aproxima das previsões do modelo complexo para a instância específica.
- O modelo simples fornece uma explicação mais fácil de entender sobre como as características da instância influenciam a previsão do modelo complexo.

Local Interpretable Model-agnostic Explanations (LIME)

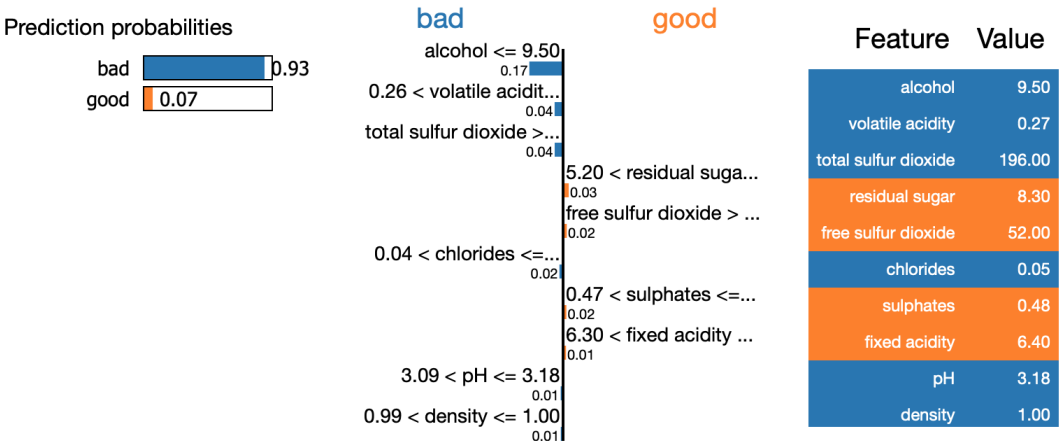
- Perturbação em variáveis categóricas é mais desafiador.
  - Perturbar variáveis categóricas requer substituir categorias por outras categorias válidas, o que é diferente de perturbar variáveis numéricas onde podemos adicionar ou subtrair valores contínuos.
- Uma alternativa é usar a codificação one-hot. Nesse caso, as perturbações são feitas em variáveis binárias.

Cor Original	Vermelho	Verde	Azul
Vermelho	1	0	0
Azul	0	0	1
Verde	0	1	0

Exemplo de codificação one-hot

Local Interpretable Model-agnostic Explanations (LIME)

Exemplo



Local Interpretable Model-agnostic Explanations (LIME)

Exemplo