

## **Aula 02**

*Banco do Brasil (Escriturário - Agente de  
Tecnologia) Passo Estratégico de  
Tecnologia de Informação - 2023  
(Pós-Edital)*

Autor:

**Thiago Rodrigues Cavalcanti**

17 de Janeiro de 2023

<b>Simulado .....</b>	<b>2</b>
<b>Questões Comentadas.....</b>	<b>9</b>
<b>Gabarito .....</b>	<b>24</b>



## SIMULADO

Vamos começar nosso primeiro simulado. Os simulados do curso do Passo Estratégico são direcionados pela análise dos últimos concursos aplicados pela banca, que no nosso caso é a Fundação Cesgranrio, e servem de treinamento para nosso concurso do **Banco do Brasil**. Esse simulado terá como base o assunto das duas primeiras aulas. Nosso objetivo é ajudar a você na fixação do assunto. Espero que goste! Qualquer dúvida estou às ordens! Forte abraço!



HORA DE  
**PRATICAR!**



LISTA DE  
**QUESTÕES**

1.

Sobre os conceitos básicos de Aprendizado de máquina assinale a alternativa correta.

- a) Teoricamente, aumentar o tamanho do conjunto de treinamento para atingir uma densidade suficiente de instâncias de treinamento não é uma solução para a maldição da dimensionalidade.
- b) No aprendizado baseado em modelo, o sistema aprende os exemplos e, em seguida, generaliza para novos casos usando uma medida de similaridade para compará-los com os exemplos aprendidos.
- c) Algoritmos de aprendizado de máquina diferentes possuem desempenho diferente mesmo que recebam uma quantidade gigantesca (suficiente) de dados.
- d) A seleção de recursos envolve a escolha dos recursos mais úteis para treinar o modelo entre os recursos existentes.
- e) Quando falamos de *overfitting* de um modelo de aprendizado, significa que o modelo tem um bom desempenho nos dados de treinamento e generaliza bem.

2.

Observe a lista de transações que aconteceram em um supermercado da sua região.

Itens	Carvão	Picanha	Maminha	Pão de Alho	Coração
T1	x	x		x	x
T2		x	x	x	x
T3	x	x		x	x
T4	x		x	x	



T5		x	x	x	x
T6	x	x		x	
T7		x	x	x	
T8		x	x	x	x
T9	x	x			x
T10	x	x		x	x

A regra de associação Carvão → Picanha tem índice de suporte igual a:

- a) 50%
- b) 83,33%
- c) 55,56%
- d) 30%
- e) 60%

3.

Qual das alternativas abaixo não apresenta um tipo de algoritmo de cluterização.

- a) Centroid-based Clustering.
- b) Density-based Clustering.
- c) Distribution-based Clustering.
- d) Hierarchical Clustering.
- e) Sklearn Clustering

4.

A regressão é um subconjunto das tarefas de aprendizagem supervisionada. Ela aprende um modelo com base em um conjunto de dados de treinamento para fazer previsões sobre dados desconhecidos ou futuros. A descrição 'supervisionado' vem do fato de que o valor de saída alvo já está definido e faz parte dos dados de treinamento. A diferença entre as subcategorias Regressão e Classificação se deve apenas ao valor de saída. Enquanto a Classificação divide o conjunto de dados em classes, a Regressão é usada para gerar valores contínuos. Qual das alternativas abaixo não pode ser visto como um método usado para regressão:

- a) Multiple Linear Regression
- b) Support Vector Regression
- c) Gaussian process regression
- d) Decision Tree



e) DBSCAN

5.

Um dos tópicos mais relevantes para aprendizado de máquina está associado aos parâmetros e hiperparâmetros dos modelos de dados. Quando dividido de acordo com essa nomenclatura parâmetro está associada aos elementos do modelo que são ajustados durante o treinamento. Já os hiperparâmetros são definidos antes do treinamento e podem melhorar a qualidade do resultado obtido. Dentro deste contexto, assinale a alternativa correta:

- a) Em um modelo de redes neurais, os pesos e vieses são considerados hiperparâmetros
- b) A quantidade de neurônios em uma camada é considerada um parâmetro do modelo de redes neurais profundas.
- c) Em um algoritmo de KNN, a métrica de distância deve ser usada para calcular a distância entre pontos é considerada um hiperparâmetro, para tal, podemos usar a distância euclidiana ou Manhattan ou ordens superiores da métrica Minkowski.
- d) Em redes neurais, as funções de ativação são usadas para introduzir uma linearidade em cada nó.
- e) A busca por hiperparâmetros ótimos é chamada de otimização de hiperparâmetros, ou seja, a busca pela combinação de hiperparâmetros para a qual o modelo treinado apresenta o melhor desempenho no conjunto de dados de treinamento.

6.

A regularização é uma técnica muito importante para reduzir o overfitting em modelos de aprendizado de máquina. Sobre esse assunto, assinale a alternativa incorreta.

- a) As normas L1 e L2 são as duas medidas de complexidade mais comuns usadas na regularização.
- b) A regularização L2 é recomendada quando nosso conjunto de dados tem muitos recursos e queremos torná-los pequenos, mas não zero.
- c) O uso da norma L1 leva à regularização L1, ou regressão Lasso.
- d) O uso da norma L2 leva à regularização L2, ou regressão Ridge.
- e) A regularização L1 é recomendada quando nosso conjunto de dados possui vários recursos e queremos transformar muitos deles em zero.

7.

A classificação é uma parte importante do aprendizado de máquina. É semelhante à regressão, pois consiste em treinar um algoritmo com dados rotulados e usá-lo para fazer previsões sobre dados futuros (não rotulados). A diferença da regressão é que, na



classificação, as previsões são categorias, como sim/não, spam/ham e assim por diante. Diante disso, assinale a alternativa correta:

- a) Os classificadores Perceptron funcionam atribuindo um peso a cada um dos recursos e um viés. A pontuação de um ponto de dados é calculada como a soma dos produtos dos pesos e características, mais o viés.
- b) A curva da característica de operação do receptor (ROC) é uma ferramenta comum usada com classificadores multiclasse.
- c) Uma maneira de avaliar o desempenho de uma regressão é observar a matriz de confusão.
- d) A descida do gradiente é um algoritmo de otimização genérico capaz de encontrar soluções ótimas para uma ampla gama de problemas. A ideia geral do algoritmo é ajustar os parâmetros iterativamente para maximizar uma função de custo.
- e) O método do gradiente descendente não se preocupa com a escala dos recursos, visto que ela não influencia na velocidade de treinamento nem no resultado do algoritmo.

8.

Se o número de recursos for alto, pode ser útil reduzi-lo com uma etapa não supervisionada antes das etapas supervisionadas. Muitos dos métodos de aprendizado não supervisionado implementam um método transform que pode ser usado para reduzir a dimensionalidade. Sobre a capacidade do scikit-learn em relação a redução de dimensionalidade assinale a alternativa correta.

- a) O PCA é usado para decompor um conjunto de dados multivariado em um conjunto de componentes não ortogonais sucessivos que explicam uma quantidade mínima da variância.
- b) No scikit-learn, o PCA é implementado como um objeto *transformer* que aprende os  $n$  componentes em seu método *predict* e pode ser usado em novos dados para projetá-los nesses componentes.
- c) Por padrão, o PCA centraliza e dimensiona os dados de entrada para cada recurso antes de aplicar o SVD.
- d) O objeto PCA fornece uma interpretação probabilística do algoritmo que pode fornecer uma probabilidade dos componentes de dados com base na quantidade de variância que ele explica.
- e) Ao plotar a variância explicada em função do número de dimensões, um método interessante para escolher os PCs é usar um gráfico de cotovelo e verificar onde a variância explicada começa diminuir.

9.

Sobre as técnicas de redução de dimensionalidade, assinale a alternativa correta:



- a) As técnicas de compressão aplicam uma codificação ou transformação para que uma representação compacta dos dados ou atributos originais seja obtida.
- b) A seleção de atributos, um dos métodos mais úteis e eficazes na compressão de dados, é um procedimento estatístico que converte um conjunto de objetos com atributos possivelmente correlacionados em um conjunto de objetos com atributos linearmente descorrelacionados, chamados de componentes principais.
- c) O número de componentes principais é maior ou igual ao número de atributos da base, e a transformação é definida de forma que o primeiro componente principal possua a menor variância.
- d) Compressão de atributos trata os valores de atributos que são substituídos por intervalos ou níveis conceituais mais elevados, reduzindo a quantidade final de atributos.
- e) Discretização efetua uma redução de dimensionalidade na qual atributos irrelevantes, pouco relevantes ou redundantes são detectados e removidos.

## 10.

Considerando as perguntas a seguir, quais delas podem ser respondidas por uma regressão.

- 1. Esse crescimento anormal de tecido é um tumor maligno?
- 2. Com base nas condições climáticas atuais, quanto vai chover amanhã?
- 3. Com base nas condições meteorológicas atuais, vai chover amanhã?
- 4. Com base no perfil de um determinado requerente, seu pedido de hipoteca deve ser aprovado?

Qual será o preço de uma determinada casa com determinadas características?

- a) 2, 3 e 4 apenas
- b) 1,4 e 5 apenas
- c) 1,2,3,4,5
- d) 2 e 5, apenas
- e) 2,4,5 apenas

## 11.

Sobre árvore de decisão no Scikit-learn assinale a alternativa correta:

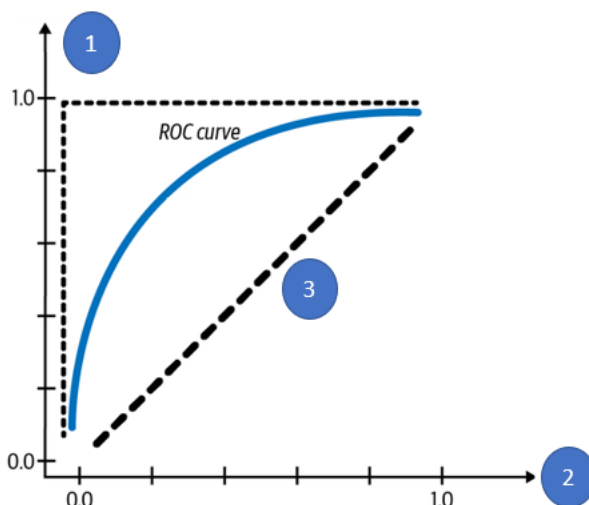
- a) O Scikit-Learn usa o Algoritmo CART, que produz apenas árvores binárias: nós não-folha sempre têm dois filhos (ou seja, as perguntas têm apenas respostas sim/não).
- b) Por padrão, a medida de impureza de entropia é usada, mas você pode selecionar a medida de gini definindo o hiperparâmetro criterion como "gini".



- c) Um dos problemas das Árvores de Decisão é que elas exigem muita preparação de dados. Elas exigem, por exemplo, dimensionamento ou centralização de recursos.
- d) As árvores de Decisão fazem várias suposições sobre os dados de treinamento.
- e) Árvore de decisão é um modelo paramétrico.

12.

A figura abaixo mostra em azul a curva ROC, além disso temos 3 círculos numerados que respectivamente se referem ao eixo das abscissas (y), eixo das ordenadas (x) e linha de base diagonal. Esses elementos são respectivamente denominados:



- a) Recall, False Alarm, Random
- b) False Alarm, Recall, Random
- c) Recall, Sensibilidade, Random
- d) Sensibilidade, Especificidade, Random
- e) Recall, Especificidade, Random

13.

A saída do seu modelo de aprendizado de máquina determina o tipo de tarefa do seu problema. Os tipos mais gerais de tarefas de ML são classificação e regressão. Dentro da classificação, existem mais subtipos, que são:

- a) Binária, Multilabel e Multiclasse





- b) De alta e baixa cardinalidade
- c) Agrupamento, regra de associação e redução de dimensionalidade
- d) CART, C4.5, ID3
- e) Árvore de decisão, SVM e Redes neurais



## QUESTÕES COMENTADAS



1.

Sobre os conceitos básicos de Aprendizado de máquina assinale a alternativa correta.

- a) Teoricamente, aumentar o tamanho do conjunto de treinamento para atingir uma densidade suficiente de instâncias de treinamento não é uma solução para a maldição da dimensionalidade.
- b) No aprendizado baseado em modelo, o sistema aprende os exemplos e, em seguida, generaliza para novos casos usando uma medida de similaridade para compará-los com os exemplos aprendidos.
- c) Algoritmos de aprendizado de máquina diferentes possuem desempenho diferente mesmo que recebam uma quantidade gigantesca (suficiente) de dados.
- d) A seleção de recursos envolve a escolha dos recursos mais úteis para treinar o modelo entre os recursos existentes.
- e) Quando falamos de *overfitting* de um modelo de aprendizado, significa que o modelo tem um bom desempenho nos dados de treinamento e generaliza bem.

**Comentários:** Vamos comentar cada uma das alternativas.

a) Em teoria, uma solução para a maldição da dimensionalidade poderia ser aumentar o tamanho do conjunto de treinamento para atingir uma densidade suficiente de instâncias de treinamento.

b) No **aprendizado baseado em instância**, o sistema aprende os exemplos de cor e, em seguida, generaliza para novos casos usando uma medida de similaridade para compará-los com os exemplos aprendidos (ou um subconjunto deles).

Outra maneira de generalizar a partir de um conjunto de exemplos é construir **um modelo** a partir desses exemplos e depois usar esse modelo para fazer previsões. Isso é chamado de **aprendizado baseado em modelo**.

c) Em um famoso artigo publicado em 2001, os pesquisadores da Microsoft Michele Banko e Eric Brill mostraram que algoritmos de aprendizado de máquina muito diferentes, incluindo os bastante simples, tiveram um desempenho quase idêntico em um problema complexo de desambiguação de linguagem natural uma vez que **receberam dados suficientes**.



d) CERTO! Uma parte crítica do sucesso de um projeto de Machine Learning é apresentar um bom conjunto de recursos para treinamento. Esse processo, chamado de **engenharia de recursos**, envolve as seguintes etapas:

**Seleção de recursos** (selecionando os recursos mais úteis para treinar entre os recursos existentes)

**Extração de recursos** (combinando recursos existentes para produzir um mais útil - como vimos anteriormente, os algoritmos de redução de dimensionalidade podem ajudar)

**Criação de recursos** (criando recursos reunindo novos dados)

e) Overfitting significa que o modelo tem um bom desempenho nos dados de treinamento, mas não generaliza bem.

**Gabarito: alternativa D.**

2.

Observe a lista de transações que aconteceram em um supermercado da sua região.

Itens	Carvão	Picanha	Maminha	Pão de Alho	Coração
T1	x	x		x	x
T2		x	x	x	x
T3	x	x		x	x
T4	x		x	x	
T5		x	x	x	x
T6	x	x		x	
T7		x	x	x	
T8		x	x	x	x
T9	x	x			x
T10	x	x		x	x

A regra de associação Carvão → Picanha tem índice de suporte igual a:

- a) 50%
- b) 83,33%
- c) 55,56%
- d) 30%
- e) 60%

**Comentário:**

O suporte trata do percentual de transações onde a regra se verifica. Neste caso, temos a transação em 5 das 10 transações. Ou seja, temos um suporte de 50%



**Gabarito: alternativa A.**

---

3.

Qual das alternativas abaixo não apresenta um tipo de algoritmo de cluterização.

- a) Centroid-based Clustering.
- b) Density-based Clustering.
- c) Distribution-based Clustering.
- d) Hierarchical Clustering.
- e) Sklearn Clustering

Comentário: Dentre as alternativas acima, a única que não representa um tipo de clustering é o sklearn clustering. O sklearn vai apresentar um conjunto de algoritmos de clustering presentes na biblioteca, mas não podem ser definidos como um tipo de algoritmo de clusterização. Os vários tipos de agrupamento são:

- Clustering baseado em conectividade (clustering hierárquico)
- Clustering baseado em centroides (métodos de particionamento)
- Clustering baseado em distribuição
- Clustering baseado em densidade (métodos baseados em modelo)
- Agrupamento difuso
- Baseado em restrições (agrupamento supervisionado)

**Gabarito: alternativa E.**

---

4.

A regressão é um subconjunto das tarefas de aprendizagem supervisionada. Ela aprende um modelo com base em um conjunto de dados de treinamento para fazer previsões sobre dados desconhecidos ou futuros. A descrição 'supervisionado' vem do fato de que o valor de saída alvo já está definido e faz parte dos dados de treinamento. A diferença entre as subcategorias Regressão e Classificação se deve apenas ao valor de saída. Enquanto a Classificação divide o conjunto de dados em classes, a Regressão é usada para gerar valores contínuos. Qual das alternativas abaixo não pode ser visto como um método usado para regressão:

- a) Multiple Linear Regression
- b) Support Vector Regression
- c) Gaussian process regression
- d) Decision Tree



### e) DBSCAN

**Comentário:** Das alternativas acima, apenas a letra E não é considerado um tipo de algoritmo de classificação. O DBSCAN (Density-Based Spatial Clustering of Applications with Noise) é um algoritmo de **clusterização** baseado em densidade semelhante ao desvio médio.

Já as demais alternativas serão definidas abaixo:

**Regressão linear múltipla** - Os modelos de regressão linear assumem que as relações entre as variáveis de entrada e de saída são lineares. Esses modelos são bastante simplistas, mas em muitos casos fornecem representações adequadas e tratáveis dos relacionamentos.

**Support Vector Regression** – A funcionalidade do Support Vector Regression (SVR) é baseada no Support Vector Machine (SVM). O objetivo da Regressão SV é encontrar uma linha reta como modelo para os pontos de dados, enquanto os parâmetros da linha reta devem ser definidos de tal forma que a linha seja o mais 'plana' possível

**Gaussian process regression** - O processo gaussiano captura o comportamento típico de um sistema com base nas observações de um sistema e fornece como resultado uma distribuição de probabilidade de possíveis funções de interpolação para o problema em questão. A Processo Gaussiano de Regressão faz uso do teorema de Bayes.

**Árvore de decisão** – é possível usar uma árvore de decisão para regressão. Neste caso usamos, por exemplo, o Mean Squared Error (MSE) como medida de impureza de um nó.

**Gabarito: alternativa E.**

## 5.

Um dos tópicos mais relevantes para aprendizado de máquina está associado aos parâmetros e hiperparâmetros dos modelos de dados. Quando dividido de acordo com essa nomenclatura parâmetro está associada aos elementos do modelo que são ajustados durante o treinamento. Já os hiperparâmetros são definidos antes do treinamento e podem melhorar a qualidade do resultado obtido. Dentro deste contexto, assinale a alternativa correta:

- a) Em um modelo de redes neurais, os pesos e vieses são considerados hiperparâmetros
- b) A quantidade de neurônios em uma camada é considerada um parâmetro do modelo de redes neurais profundas.
- c) Em um algoritmo de KNN, a métrica de distância deve ser usada para calcular a distância entre pontos é considerada um hiperparâmetro, para tal, podemos usar a distância euclidiana ou Manhattan ou ordens superiores da métrica Minkowski.
- d) Em redes neurais, as funções de ativação são usadas para introduzir uma linearidade em cada nó.
- e) A busca por hiperparâmetros ótimos é chamada de otimização de hiperparâmetros, ou seja, a busca pela combinação de hiperparâmetros para a qual o modelo treinado apresenta o melhor desempenho no conjunto de dados de treinamento.

**Comentário:** Vamos comentar cada uma das alternativas:



- a) Os pesos e vieses são **parâmetros** associados aos modelos de redes neurais. Durante o treinamento, esses valores são ajustados como o objetivo de melhorar a capacidade de previsão do modelos. Neste caso, uma função custo é definida e a nossa meta é atingir o menor valor possível para esta função fazendo os ajustes nos parâmetros.
- b) A quantidade de neurônios por camadas é considerada um **hiperparâmetro** do modelo de redes neurais.
- c) O algoritmo K-nearest neighbor (KNN) pode ser usado como um algoritmo de aprendizado de máquina supervisionado ou não supervisionado e pode ser aplicado a problemas de classificação, regressão, agrupamento e detecção de valores discrepantes. Em um algoritmo de KNN, a métrica de distância deve ser usada para calcular a distância entre pontos é considerada um hiperparâmetro, para tal, podemos usar a distância euclidiana ou Manhattan ou ordens superiores da métrica Minkowski.
- d) As funções de ativação são usadas para introduzir uma **não linearidade** em cada nó. Poucas coisas que precisamos ter certeza ao decidir as funções de ativação são, elas devem ser usadas em milhares e milhões de nós, e a propagação reversa usa suas derivadas, então tanto a função quanto sua derivada devem ser menos complexas computacionalmente. Algumas das ativações amplamente utilizadas são ReLU, Sigmoid e Leaky ReLU.
- e) A busca por hiperparâmetros ótimos é chamada de otimização de hiperparâmetros, ou seja, a busca pela combinação de hiperparâmetros para a qual o modelo treinado apresenta o melhor desempenho para um determinado conjunto de dados (generalização).

**Gabarito: alternativa C.**

6.

A regularização é uma técnica muito importante para reduzir o *overfitting* em modelos de aprendizado de máquina. Sobre esse assunto, assinale a alternativa incorreta.

- a) As normas L1 e L2 são as duas medidas de complexidade mais comuns usadas na regularização.
- b) A regularização L2 é recomendada quando nosso conjunto de dados tem muitos recursos e queremos torná-los pequenos, mas não zero.
- c) O uso da norma L1 leva à regularização L1, ou regressão Lasso.
- d) O uso da norma L2 leva à regularização L2, ou regressão Ridge.
- e) A regularização L1 é recomendada quando nosso conjunto de dados possui vários recursos e queremos transformar muitos deles em zero.

**Comentário:** A regularização é uma técnica muito importante para reduzir o *overfitting* em modelos de aprendizado de máquina. Consiste em adicionar uma medida de complexidade (termo de regularização) à função de erro durante o processo de treinamento.

As normas L1 e L2 são as duas medidas de complexidade mais comuns usadas na regularização.



O uso da norma L1 leva à regularização L1, ou regressão de laço. O uso da norma L2 leva à regularização L2, ou regressão da crista.

A regularização L1 é recomendada quando nosso conjunto de dados possui vários recursos e queremos transformar muitos deles em zero. A regularização L2 é recomendada quando nosso conjunto de dados tem **poucos recursos** e queremos torná-los pequenos, mas não zero.

Assim, observamos que a alternativas B está errada.

**Gabarito: alternativa B.**

7.

A classificação é uma parte importante do aprendizado de máquina. É semelhante à regressão, pois consiste em treinar um algoritmo com dados rotulados e usá-lo para fazer previsões sobre dados futuros (não rotulados). A diferença da regressão é que, na classificação, as previsões são categorias, como sim/não, spam/ham e assim por diante. Diante disso, assinale a alternativa correta:

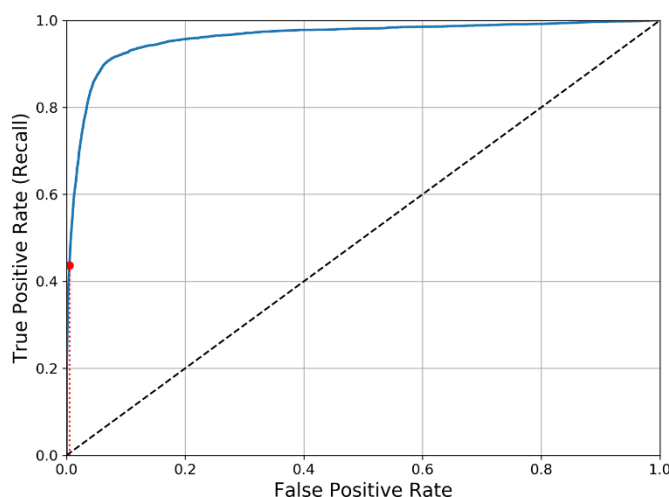
- a) Os classificadores Perceptron funcionam atribuindo um peso a cada um dos recursos e um viés. A pontuação de um ponto de dados é calculada como a soma dos produtos dos pesos e características, mais o viés.
- b) A curva da característica de operação do receptor (ROC) é uma ferramenta comum usada com classificadores multiclasse.
- c) Uma maneira de avaliar o desempenho de uma regressão é observar a matriz de confusão.
- d) A descida do gradiente é um algoritmo de otimização genérico capaz de encontrar soluções ótimas para uma ampla gama de problemas. A ideia geral do algoritmo é ajustar os parâmetros iterativamente para maximizar uma função de custo.
- e) O método do gradiente descendente não se preocupa com a escala dos recursos, visto que ela não influencia na velocidade de treinamento nem no resultado do algoritmo.

**Comentário:** Vejamos os erros das demais alternativas:

- b) A curva da característica de operação do receptor (ROC) é uma ferramenta comum usada com classificadores **binários**. A curva ROC plota a taxa de verdadeiros positivos (outro nome para recall) em relação à taxa de falsos positivos (FPR).



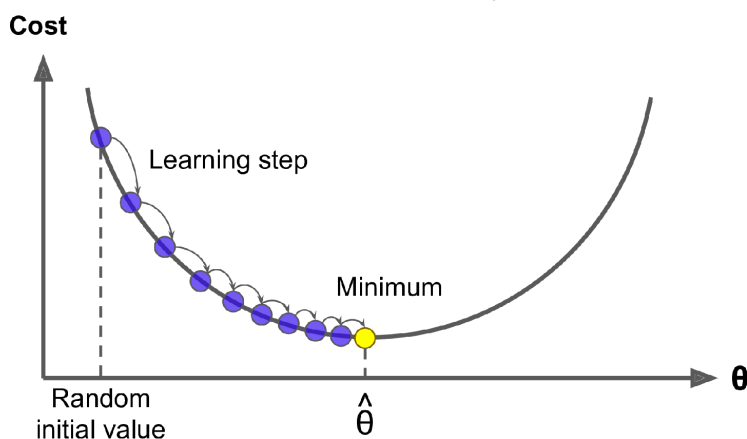




A linha pontilhada representa a curva ROC de um classificador puramente aleatório; um bom classificador fica o mais longe possível dessa linha (em direção ao canto superior esquerdo).

c) Uma maneira de avaliar o desempenho de **um classificador** é observar a matriz de confusão.

d) A descida do gradiente é um algoritmo de otimização genérico capaz de encontrar soluções ótimas para uma ampla gama de problemas. A ideia geral do gradiente descendente é ajustar os parâmetros iterativamente para **minimizar** uma função de custo.



Nesta representação de Gradient Descent, os parâmetros do modelo são inicializados aleatoriamente e ajustados repetidamente para minimizar a função de custo; o tamanho da etapa de aprendizado é proporcional à inclinação da função de custo, de modo que as etapas diminuam gradualmente à medida que os parâmetros se aproximam do mínimo

e) Ao usar o Gradiente Descendente, você deve garantir que todos os recursos tenham uma escala semelhante (por exemplo, usando a classe StandardScaler do Scikit-Learn), ou levará muito mais tempo para convergir.

**Gabarito: alternativa A.**



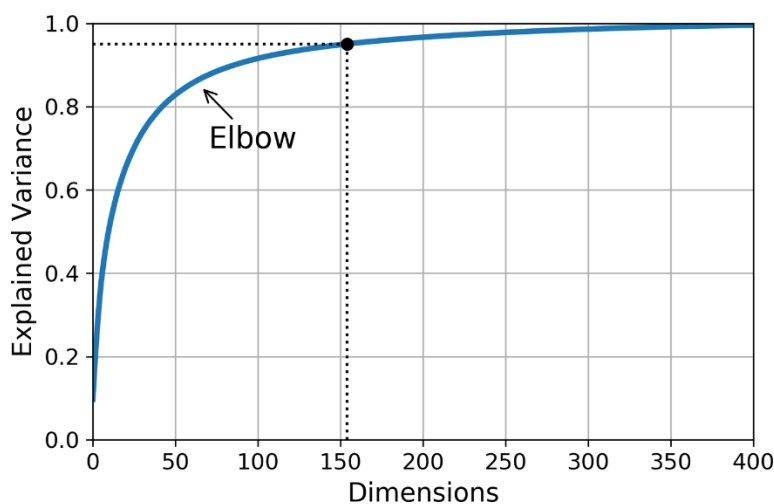
Se o número de recursos for alto, pode ser útil reduzi-lo com uma etapa não supervisionada antes das etapas supervisionadas. Muitos dos métodos de aprendizado não supervisionado implementam um método transform que pode ser usado para reduzir a dimensionalidade. Sobre a capacidade do scikit-learn em relação a redução de dimensionalidade assinale a alternativa correta.

- a) O PCA é usado para decompor um conjunto de dados multivariado em um conjunto de componentes não ortogonais sucessivos que explicam uma quantidade mínima da variância.
- b) No scikit-learn, o PCA é implementado como um objeto *transformer* que aprende os *n* componentes em seu método *predict* e pode ser usado em novos dados para projetá-los nesses componentes.
- c) Por padrão, o PCA centraliza e dimensiona os dados de entrada para cada recurso antes de aplicar o SVD.
- d) O objeto PCA fornece uma interpretação probabilística do algoritmo que pode fornecer uma probabilidade dos componentes de dados com base na quantidade de variância que ele explica.
- e) Ao plotar a variância explicada em função do número de dimensões, um método interessante para escolher os PCs é usar um gráfico de cotovelo e verificar onde a variância explicada começa diminuir.

**Comentário:** Vamos comentar cada uma das alternativas:

- a) **Errado.** O PCA é usado para decompor um conjunto de dados multivariado em um conjunto de componentes **ortogonais** sucessivos que explicam uma quantidade mínima da variância.
- b) **Errado.** No scikit-learn, o PCA é implementado como um objeto *transformer* que aprende os *n* componentes em seu método **fit** e pode ser usado em novos dados para projetá-los nesses componentes.
- c) **Errado.** O PCA centraliza, mas não dimensiona os dados de entrada para cada recurso antes de aplicar o SVD. O parâmetro opcional `whiten=True` torna possível projetar os dados no espaço singular enquanto dimensiona cada componente para a variação da unidade. Isso geralmente é útil se os modelos a jusante fizerem suposições fortes sobre a isotropia do sinal: esse é, por exemplo, o caso de máquinas de vetor de suporte com o kernel RBF e o algoritmo de agrupamento K-Means.
- d) **Certo.** A razão de variância explicada de cada componente principal é disponibilizada através da variável `explained_variance_ratio_`. A razão indica a proporção da variância do conjunto de dados que se encontra ao longo de cada componente principal.
- e) **Errado.** Ao plotar a variância explicada em função do número de dimensões, um método interessante para escolher os PCs é usar um gráfico de cotovelo e verificar onde a variância explicada começa **para de crescer rapidamente**.





Variação explicada em função do número de dimensões

**Gabarito: alternativa D.**

9.

Sobre as técnicas de redução de dimensionalidade, assinale a alternativa correta:

- a) As técnicas de compressão aplicam uma codificação ou transformação para que uma representação compacta dos dados ou atributos originais seja obtida.
- b) A seleção de atributos, um dos métodos mais úteis e eficazes na compressão de dados, é um procedimento estatístico que converte um conjunto de objetos com atributos possivelmente correlacionados em um conjunto de objetos com atributos linearmente decorrelacionados, chamados de componentes principais.
- c) O número de componentes principais é maior ou igual ao número de atributos da base, e a transformação é definida de forma que o primeiro componente principal possua a menor variância.
- d) Compressão de atributos trata os valores de atributos que são substituídos por intervalos ou níveis conceituais mais elevados, reduzindo a quantidade final de atributos.
- e) Discretização efetua uma redução de dimensionalidade na qual atributos irrelevantes, pouco relevantes ou redundantes são detectados e removidos.

**Comentários:** Vamos comentar cada uma das alternativas:

- a) (CERTO) As técnicas de compressão aplicam uma codificação ou transformação para que uma representação compacta dos dados ou atributos originais seja obtida.
- b) (ERRADO) A análise de componentes principais (Principal Component Analysis – PCA), um dos métodos mais úteis e eficazes na compressão de dados, é um procedimento estatístico que converte um conjunto de objetos com atributos possivelmente correlacionados em um conjunto de objetos com atributos linearmente decorrelacionados, chamados de componentes principais.



- c) (ERRADO) O número de componentes principais é menor ou igual ao número de atributos da base, e a transformação é definida de forma que o primeiro componente principal possua a maior variância.
- d) (ERRADO) Discretização trata os valores de atributos que são substituídos por intervalos ou níveis conceituais mais elevados, reduzindo a quantidade final de atributos.
- e) (ERRADO) Seleção de atributos (ou características) efetua uma redução de dimensionalidade na qual atributos irrelevantes, pouco relevantes ou redundantes são detectados e removidos.

Assim, temos a nossa resposta na alternativa A.

**Gabarito: alternativa A.**

10.

Considerando as perguntas a seguir, quais delas podem ser respondidas por uma regressão.

1. Esse crescimento anormal de tecido é um tumor maligno?
2. Com base nas condições climáticas atuais, quanto vai chover amanhã?
3. Com base nas condições meteorológicas atuais, vai chover amanhã?
4. Com base no perfil de um determinado requerente, seu pedido de hipoteca deve ser aprovado?

Qual será o preço de uma determinada casa com determinadas características?

- a) 2, 3 e 4 apenas
- b) 1,4 e 5 apenas
- c) 1,2,3,4,5
- d) 2 e 5, apenas
- e) 2,4,5 apenas

**Comentários:** O tipo de variável de destino determina o tipo de modelo de aprendizado de máquina supervisionado que temos. Basicamente, temos dois tipos de modelos de aprendizado de máquina supervisionados:

**Classificadores:** se a variável de destino for uma variável categórica, o modelo de aprendizado de máquina é chamado de **classificador**. Os classificadores podem ser usados para responder aos seguintes tipos de perguntas de negócios:

- Esse crescimento anormal de tecido é um tumor maligno?
- Com base nas condições meteorológicas atuais, vai chover amanhã?
- Com base no perfil de um determinado requerente, seu pedido de hipoteca deve ser aprovado?



**Regressores:** se a variável de destino é uma **variável contínua**, treinamos um regressor. Os regressores podem ser usados para responder aos seguintes tipos de perguntas de negócios:

- Com base nas condições climáticas atuais, quanto vai chover amanhã? [2]
- Qual será o preço de uma determinada casa com determinadas características? [5]

Assim, temos a nossa resposta na alternativa D

**Gabarito: alternativa D.**

11.

Sobre árvore de decisão no Scikit-learn assinale a alternativa correta:

- a) O Scikit-Learn usa o Algoritmo CART, que produz apenas árvores binárias: nós não-folha sempre têm dois filhos (ou seja, as perguntas têm apenas respostas sim/não).
- b) Por padrão, a medida de impureza de entropia é usada, mas você pode selecionar a medida de gini definindo o hiperparâmetro criterion como "gini".
- c) Um dos problemas das Árvores de Decisão é que elas exigem muita preparação de dados. Elas exigem, por exemplo, dimensionamento ou centralização de recursos.
- d) As árvores de Decisão fazem várias suposições sobre os dados de treinamento.
- e) Árvore de decisão é um modelo paramétrico.

**Comentário:** Vamos comentar cada uma das alternativas:

- a) O Scikit-Learn usa o Algoritmo CART, que produz apenas árvores binárias: nós não-folha sempre têm dois filhos (ou seja, as perguntas têm apenas respostas sim/não). No entanto, outros algoritmos como o ID3 podem produzir Árvores de Decisão com nós que possuem mais de dois filhos.
- b) Por padrão, a medida de impureza Gini é usada, mas você pode selecionar a medida de impureza de entropia definindo o hiperparâmetro criterion como "entropy".
- c) Uma das muitas qualidades das Árvores de Decisão é que elas exigem muito pouca preparação de dados. Na verdade, elas não exigem dimensionamento ou centralização de recursos.
- d) As árvores de Decisão fazem muito poucas suposições sobre os dados de treinamento (em oposição aos modelos lineares, que assumem que os dados são lineares, por exemplo). Se não houver restrições, a estrutura da árvore se adaptará aos dados de treinamento, ajustando-os muito bem - na verdade, provavelmente superajustando-os.
- e) Um tal modelo de árvore de decisão é frequentemente chamado de modelo não paramétrico, não porque não tenha nenhum parâmetro (geralmente tem muitos), mas porque o número de parâmetros não é determinado antes do treinamento, de modo que a estrutura do modelo é livre para se ater aos dados. Em contraste, um modelo paramétrico, como um modelo linear, possui um número predeterminado de parâmetros, de modo que

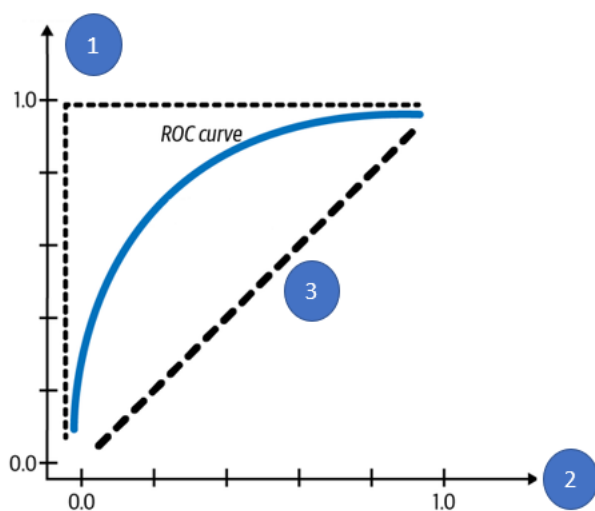


seu grau de liberdade é limitado, reduzindo o risco de overfitting (mas aumentando o risco de underfitting).

**Gabarito: alternativa A.**

12.

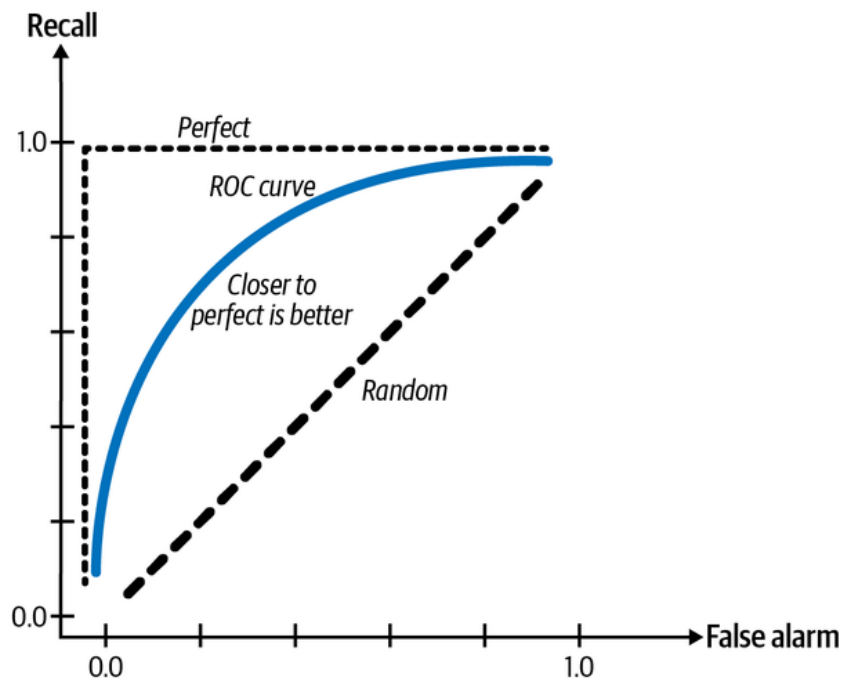
A figura abaixo mostra em azul a curva ROC, além disso temos 3 círculos numerados que respectivamente se referem ao eixo das abscissas (y), eixo das ordenadas (x) e linha de base diagonal. Esses elementos são respectivamente denominados:



- a) Recall, False Alarm, Random
- b) False Alarm, Recall, Random
- c) Recall, Sensibilidade, Random
- d) Sensibilidade, Especificidade, Random
- e) Recall, Especificidade, Random

Comentário:





Muitos problemas de classificação podem ser modelados como problemas de regressão. Seu modelo pode gerar uma probabilidade e, com base nessa probabilidade, você classifica a amostra. Por exemplo, se o valor for maior que 0,5, é um rótulo positivo e, se for menor ou igual a 0,5, é um rótulo negativo. Isso significa que você pode ajustar o limite para aumentar a *taxa de verdadeiros positivos* (também conhecida como *recall*) enquanto diminui a *taxa de falsos positivos* (também conhecida como *probabilidade de alarme falso*) e vice-versa. Podemos plotar a taxa de verdadeiros positivos em relação à taxa de falsos positivos para diferentes limites. Este gráfico é conhecido como *curva ROC* (*receiver operating characteristic*). Quando seu modelo é perfeito, o *recall* é 1,0, e a curva é apenas uma linha no topo. Essa curva mostra como o desempenho do seu modelo muda dependendo do limite e ajuda a escolher o limite que funciona melhor para você. Quanto mais próximo da linha perfeita, melhor o desempenho do seu modelo.

**Gabarito: alternativa A.**

13.

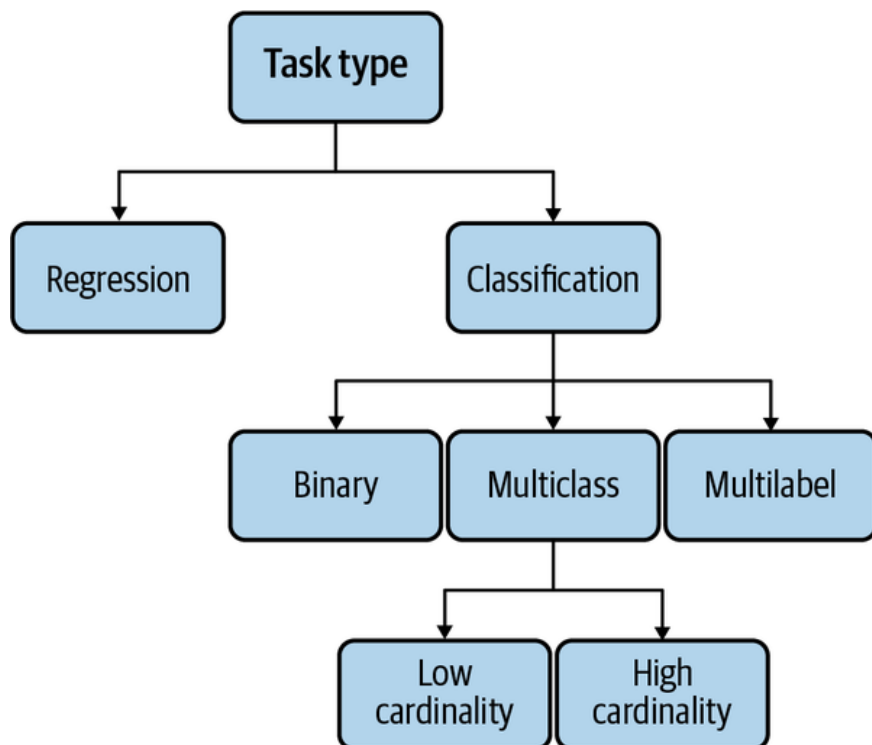
A saída do seu modelo de aprendizado de máquina determina o tipo de tarefa do seu problema. Os tipos mais gerais de tarefas de ML são classificação e regressão. Dentro da classificação, existem mais subtipos, que são:

- a) Binária, Multilabel e Multiclasse
- b) De alta e baixa cardinalidade
- c) Agrupamento, regra de associação e redução de dimensionalidade
- d) CART, C4.5, ID3
- e) Árvore de decisão, SVM e Redes neurais



Comentário:

A figura abaixo mostra os subtipos de tarefas de classificação:



Nos problemas de classificação, quanto menos classes houver para classificar, mais simples será o problema. A mais simples é a **classificação binária**, onde existem apenas duas classes possíveis. Exemplos de classificação binária incluem classificar se um comentário é tóxico, se uma varredura pulmonar mostra sinais de câncer, se uma transação é fraudulenta.

Quando há mais de duas classes, o problema passa a ser **classificação multiclasse**. Lidar com problemas de classificação binária é muito mais fácil do que lidar com problemas multiclasse.

Quando o número de classes é alto, como diagnóstico de doenças onde o número de doenças pode chegar a milhares ou classificações de produtos onde o número de produtos pode chegar a dezenas de milhares, dizemos a tarefa de classificação tem **alta cardinalidade**. Problemas de alta cardinalidade podem ser muito desafiadores. O primeiro desafio está na coleta de dados. A coleta de dados pode ser especialmente difícil para classes raras. Quando você tem milhares de classes, é provável que algumas delas sejam raras.

Tanto na classificação binária quanto na multiclasse, cada exemplo pertence a exatamente uma classe. Quando um exemplo pode pertencer a várias classes, temos um problema de **classificação multirrótulo**. Por exemplo, ao construir um modelo para classificar artigos em quatro tópicos – tecnologia, entretenimento, finanças e política – um artigo pode ser de tecnologia e finanças.

Existem duas abordagens principais para problemas de classificação multirrótulo. A primeira é tratá-la como se fosse uma classificação multiclasse. Na classificação multiclasse, se houver quatro classes possíveis [tecnologia, entretenimento, finanças, política] e o rótulo por exemplo for entretenimento, você representa esse rótulo com o vetor [0, 1, 0, 0]. Na classificação





multilabel, se um exemplo tiver os rótulos entretenimento e finanças, seu rótulo será representado como [0, 1, 1, 0].

A segunda abordagem é transformá-lo em um conjunto de problemas de classificação binária. Para o problema de classificação de artigos, você pode ter quatro modelos correspondentes a quatro tópicos, cada modelo informando se um artigo está nesse tópico ou não.

**Gabarito: alternativa A.**

---





## GABARITO

1. D
2. A
3. E
4. E
5. C
6. B
7. A
8. D
9. A
10. D
11. A
12. A
13. A

Forte abraço e bons estudos.

**"Hoje, o 'Eu não sei', se tornou o 'Eu ainda não sei'"**

(Bill Gates)

# Thiago Cavalcanti



**You**Tube

**Face:** [www.facebook.com/profthiagocavalcanti](https://www.facebook.com/profthiagocavalcanti)

**Insta:** [www.instagram.com/prof.thiago.cavalcanti](https://www.instagram.com/prof.thiago.cavalcanti)

**YouTube:** [youtube.com/profthiagocavalcanti](https://youtube.com/profthiagocavalcanti)



# ESSA LEI TODO MUNDO CONHECE: PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.



1 Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.



2 Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).



3 Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).



4 Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).



5 Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).



6 Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).



7 Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.



8 O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.



Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.