



Aprendizado de Máquina

Classificação

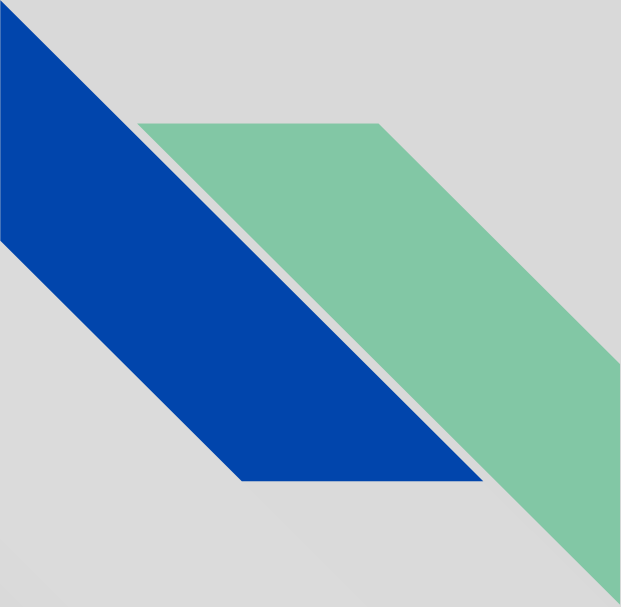
Por Gustavo Alexandre

“Os erros causados por dados inadequados são muito menores do que aqueles devido à falta total de dados” *Charles Babbage*



Agenda

- 1. Mercado**
- 2. Conceitos**
- 3. Dados**
- 4. Classificação**
- 5. Avaliação**
- 6. Técnicas**
- 7. Considerações**



Mercado

"Há três tipos de mentiras: as mentiras, as mentiras descabeladas, e as estatísticas"

Benjamin Disraeli

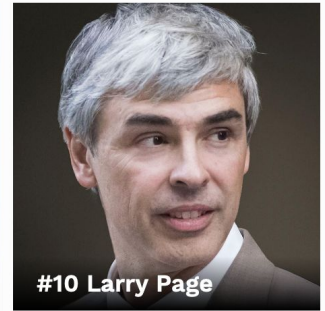
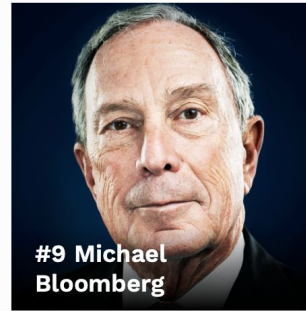
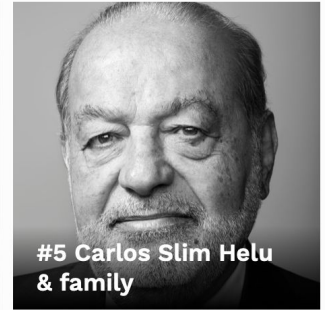
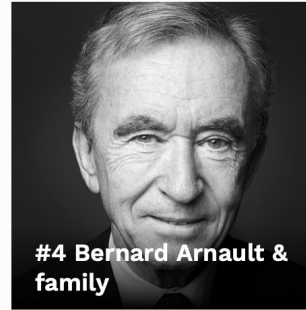


Mercado

Segundo a Forbes, "*Das 25 maiores empresas de tecnologia, 15 são norte-americanas, com oito delas entre as Top 10: Apple, Samsung, Microsoft, Alphabet, IBM, Intel, Cisco Systems, Oracle, Facebook e Hon Hai Precision Industry.*

"

Mercado



Mercado



Mercado



Mercado



Mercado



Biggest Tech Company Ever

Alphabet

Inc.



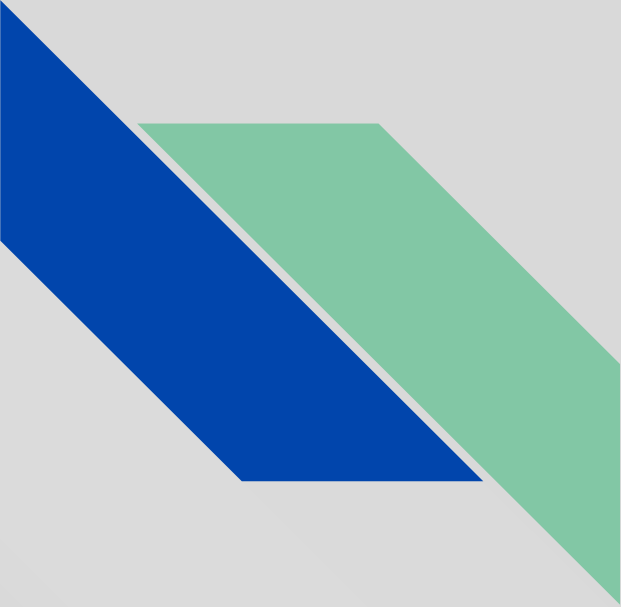
is for Google

Mercado





Transformação Digital



Conceitos

"Há três tipos de mentiras: as mentiras, as mentiras descabeladas, e as estatísticas"

Benjamin Disraeli



Conceitos

O Aprendizado de Máquina (AM) é uma subárea da inteligência artificial que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional

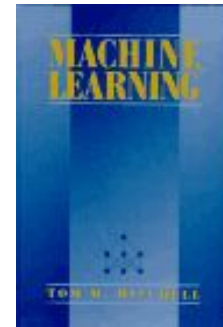
Conceitos

- Arthur Samuel (1959): “área de estudo que concede aos computadores a habilidade de aprender sem serem programados explicitamente.” [[samuel1959](#)]



Conceitos

Tom Mitchell (1998): “*Um programa de computador aprende com a experiência E em relação a tarefa T e alguma medida de desempenho P , se seu desempenho em T , medido por P , melhora com a experiência E .*”
[[Mitchell1998](#)]





Conceitos

O Aprendizado de Máquina explora o estudo e a construção de algoritmos que podem aprender sobre dados e fazer previsões



Aprendizado de Máquina

- **Supervisionado** visa construir um modelo estatístico a partir de um conjunto de dados que contém as entradas e as saídas desejadas (rotuladas).
 - Classificação
 - Regressão



Conceitos

- **Não-Supervisionado** visa construir um modelo estatístico a partir de um conjunto de dados que contém apenas entradas e nenhum rótulo de saída desejado (não-rotulada).
 - Agrupamento
 - Sistemas de Recomendação
 - Filtragem



Conceitos

- **Semi-Supervisionado** compreende o desenvolvimento de modelos matemáticos a partir dos dados de treinamento incompletos, em que uma parte da entrada de amostra não possui rótulos (rotulados e não-rotulados).



Conceitos

- **Profundo** compreende o uso das redes neurais artificiais em grandes quantidades de dados, ampliando continuamente sua capacidade de aprendizado, à medida que processam mais dados.



Conceitos

- **Por Reforço** compreende a investigação sobre a forma como agentes inteligentes devem agir em determinados ambientes, de modo a maximizar alguma noção de recompensa cumulativa.



Aprendizado Supervisionado

"Há três tipos de mentiras: as mentiras, as mentiras descabeladas, e as estatísticas"
Benjamin Disraeli



Aprendizado Supervisionado

- A máquina recebe as respostas corretas.
- Dois tipos (tarefas):
 - **Classificação**: prediz valor discreto.
 - **Regressão**: prediz valor contínuo.

Exemplos

Detecção de spam

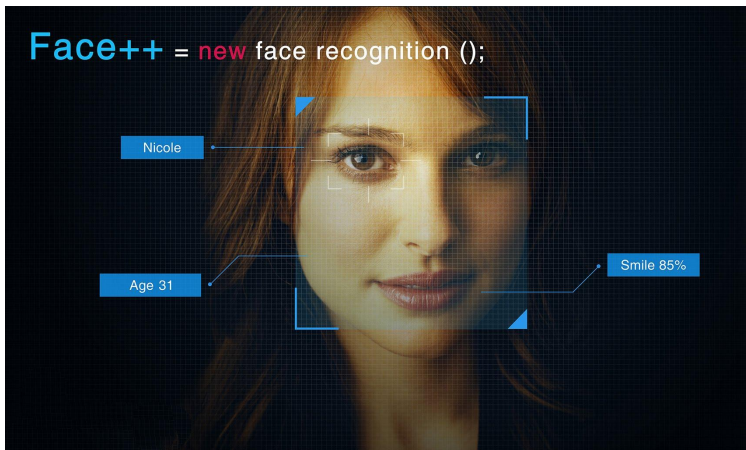


Reconhecimento de Voz

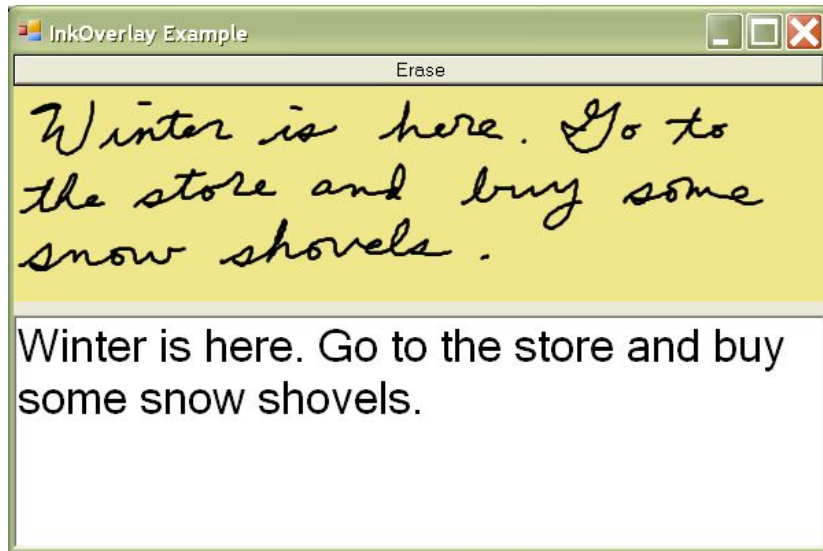


Exemplos

Reconhecimento de imagens



Reconhecimento de caracteres

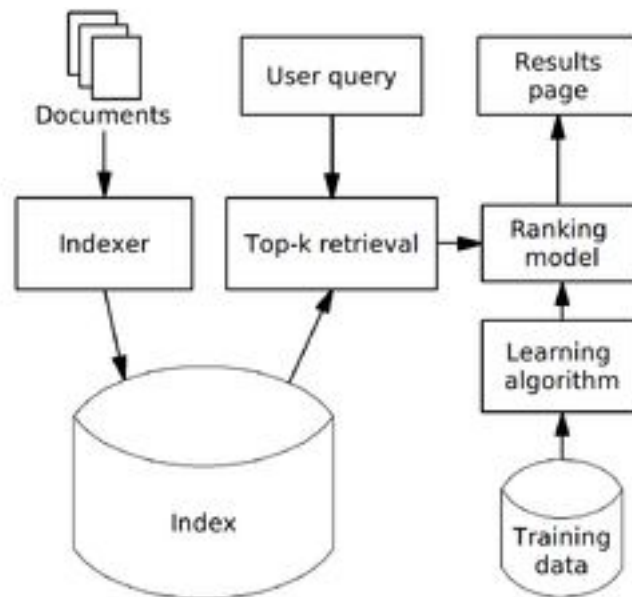


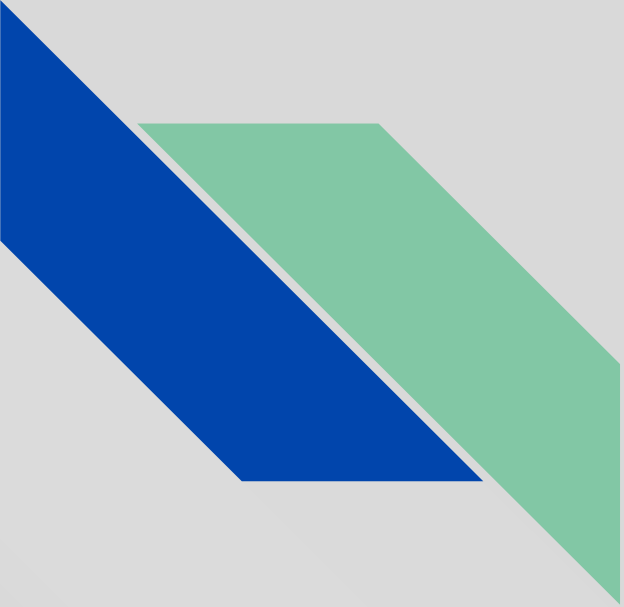
Exemplos

Tradução Automática



Learning to rank



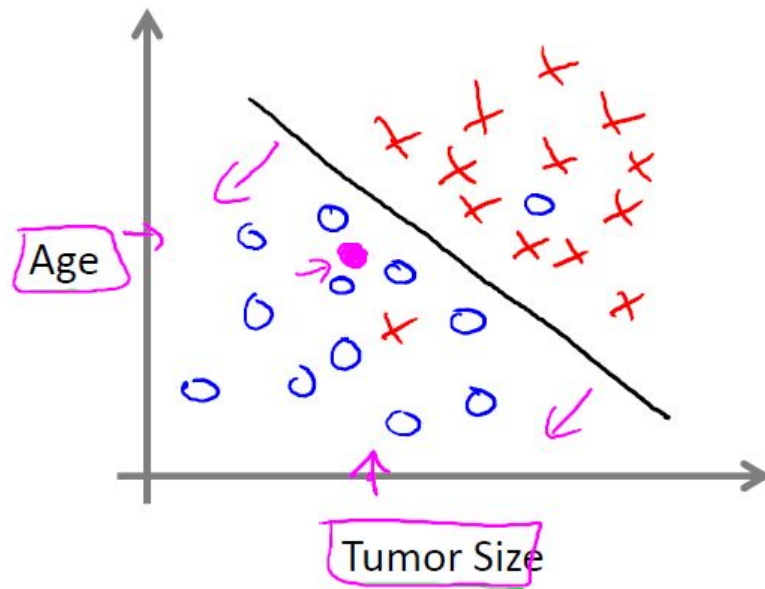


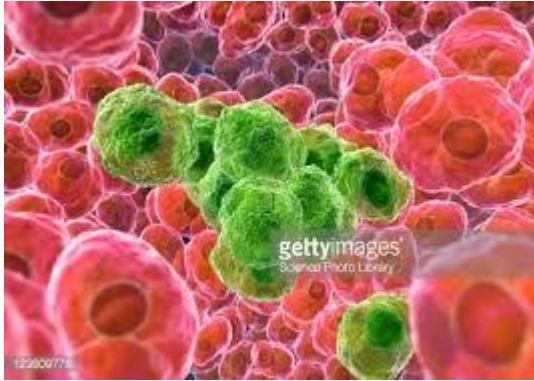
Problemas de AM



Identifique o problema de AM

- ❖ Dado um exemplo de tumor cancerígeno, temos de **prever se ele é benigno ou maligno através do seu tamanho e idade do paciente.**





- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- ...

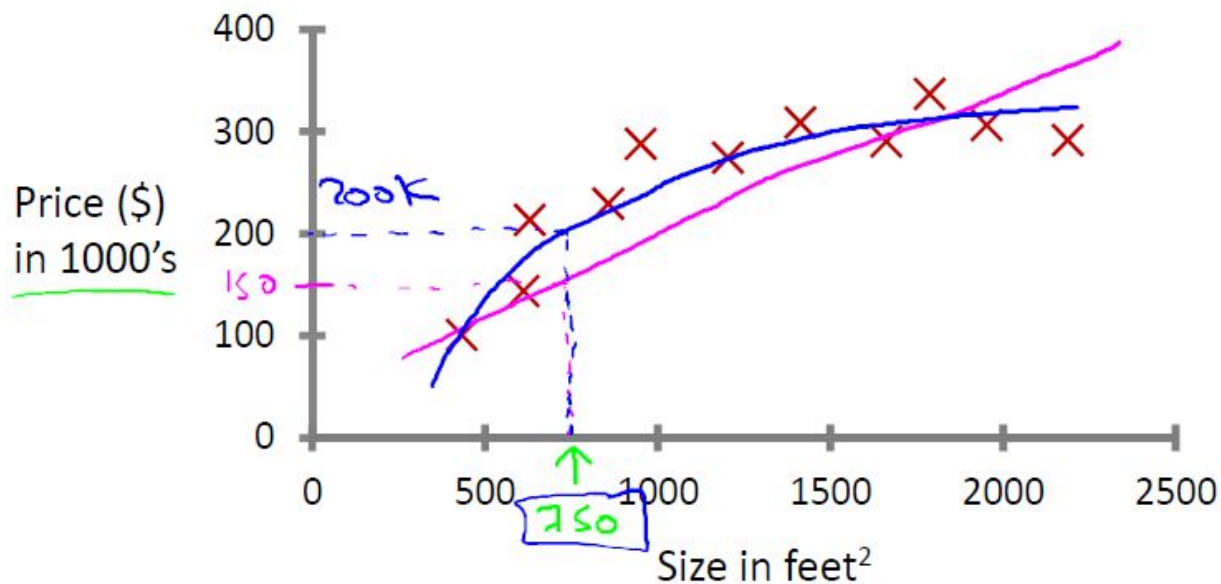


Identifique o problema de AM

- ❖ Dado um conjunto de dados sobre o tamanho de casas no mercado imobiliário, tentar **prever o preço de casas**, já que algumas instâncias foram atribuídas como padrão A, B e C.

Exemplos

Housing price prediction.





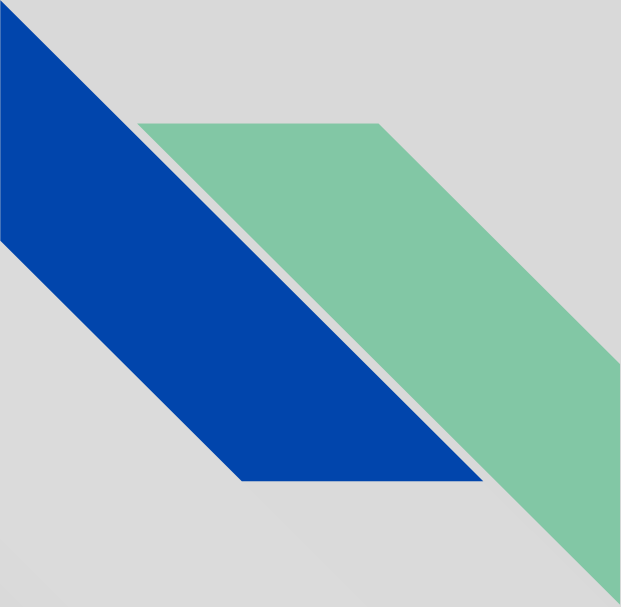
Identifique o problema de AM

- ❖ Dada uma imagem de homem ou mulher, temos de **prever sua idade com base em dados da imagem.**



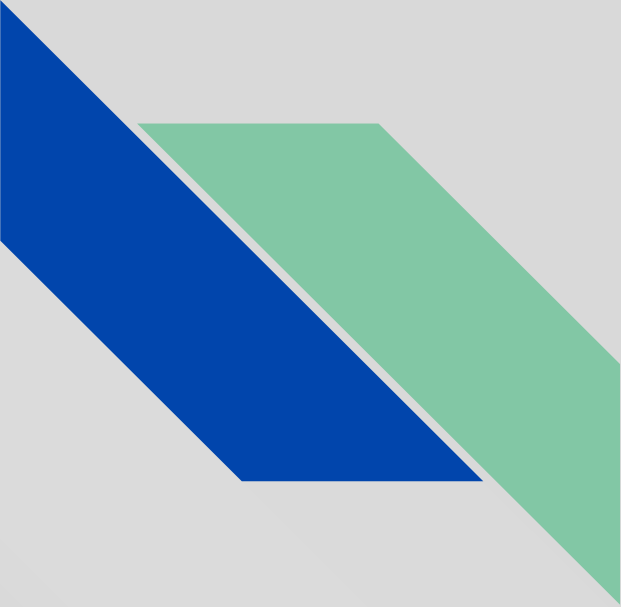
Identifique o problema de AM

- ❖ Dada uma coleção de 1000 pesquisas de uma universidade encontrar uma maneira de **agrupar automaticamente estas pesquisas em grupos que são de alguma forma semelhantes** ou relacionados por diferentes variáveis, tais como a frequência das palavras, frases e contagem de páginas.



Classificação

"No futuro, o pensamento estatístico será tão necessário para a cidadania eficiente como saber ler e escrever" *H.G.Wells*



Dados

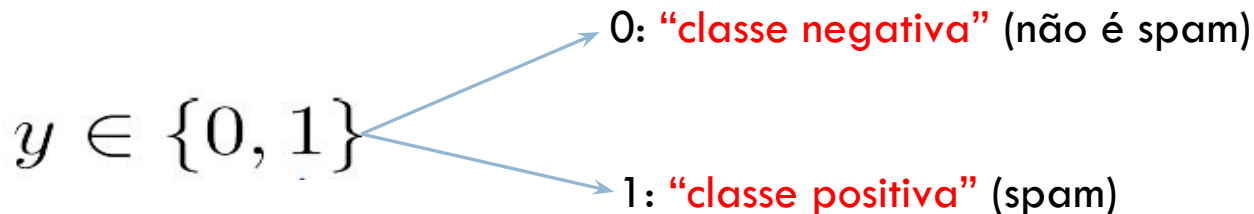
"No futuro, o pensamento estatístico será tão necessário para a cidadania eficiente como saber ler e escrever" *H.G.Wells*



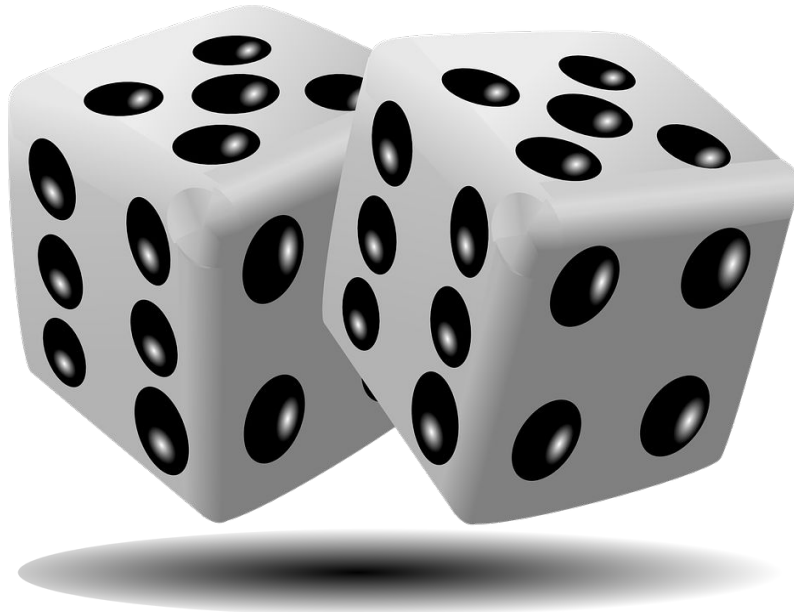
Classificação binária

● Exemplos

- **Email:** spam/ham (not spam)?
- **Transações financeiras:** fraudulenta/legítima?
- **Tumor:** maligno/benigno?



Dados

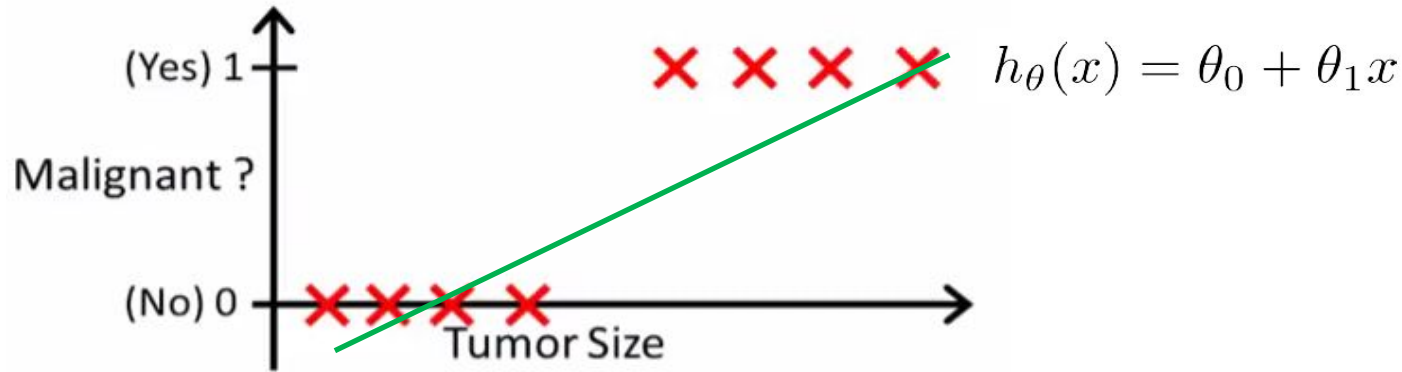




Terminologia

- ❖ Conjunto de Treinamento
- ❖ Conjunto de Teste
- ❖ Função alvo
- ❖ Hipótese
- ❖ Modelo
- ❖ Algoritmo de Aprendizado

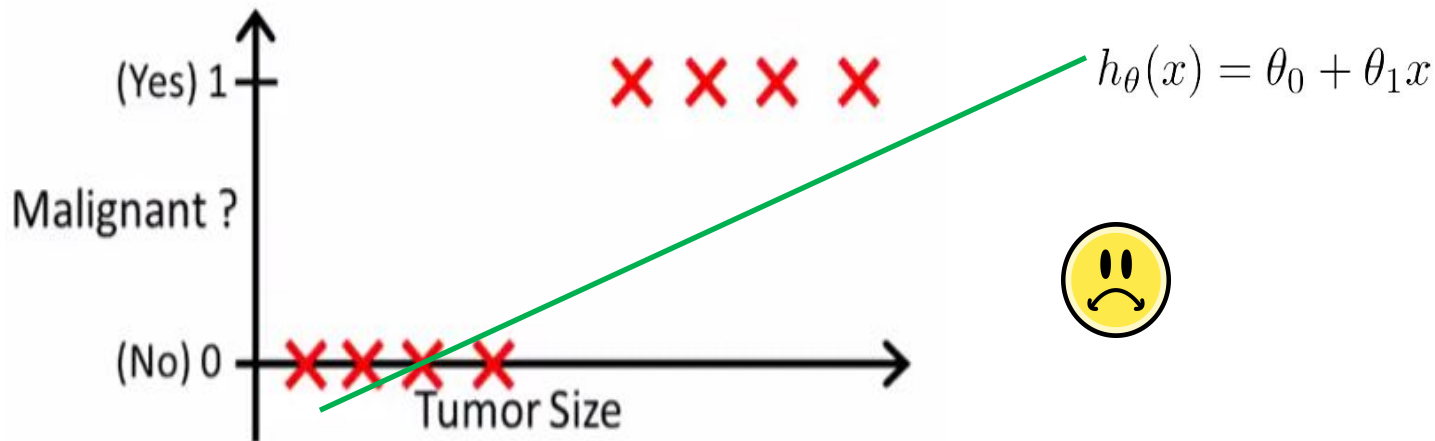
Podemos usar Regressão Linear?



If $h_{\theta}(x) \geq 0.5$, predict “y = 1”

If $h_{\theta}(x) < 0.5$, predict “y = 0”

Não podemos usar Regressão Linear!





Não podemos usar Regressão Linear!

- Em problemas classificação binária: $y=0$ ou $y=1$
- Na Regressão Linear, é possível que $h_{\theta}(x) > 1$ ou que $h_{\theta}(x) < 0$.
- Na Regressão Logística vamos definir a representação de hipóteses de forma que :

$$0 \leq h_{\theta}(x) \leq 1$$



Interpretação

$h_{\theta}(x)$: estimativa da probabilidade de que $y = 1$ para a entrada x

- Exemplo: $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$ $h_{\theta}(x) = 0.7$

Há 70% de chance de que o tumor seja maligno.

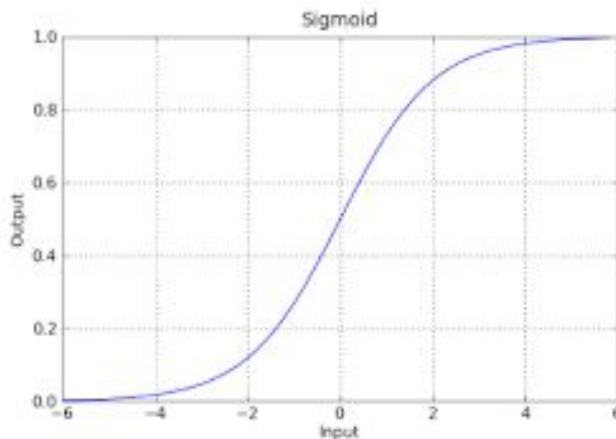
$$h_{\theta}(x) = \Pr(y = 1 \mid x; \theta) \qquad \Pr(y = 0 \mid x; \theta) = 1 - \Pr(y = 1 \mid x; \theta)$$

Fronteira de Decisão

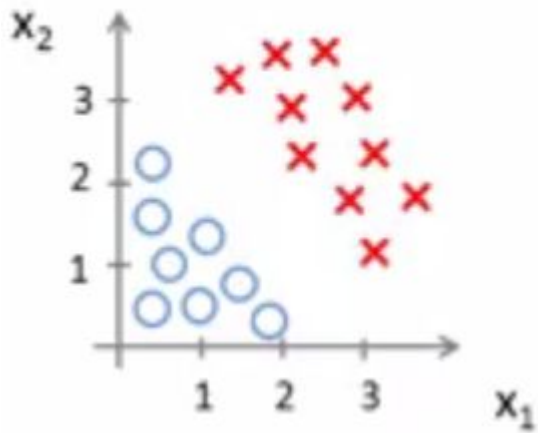
- Repare que
 $g(z) \geq 0.5$ quando $z \geq 0$
 $g(z) < 0.5$ quando $z < 0$
- Sendo assim
prediz $y=1$ quando
prediz $y=0$ quando

$$h_{\theta}(x) = g(\theta^T x) = \Pr(y = 1 \mid x; \theta)$$

$$g(z) = \frac{1}{1 + e^{-z}}$$



Fronteira Linear

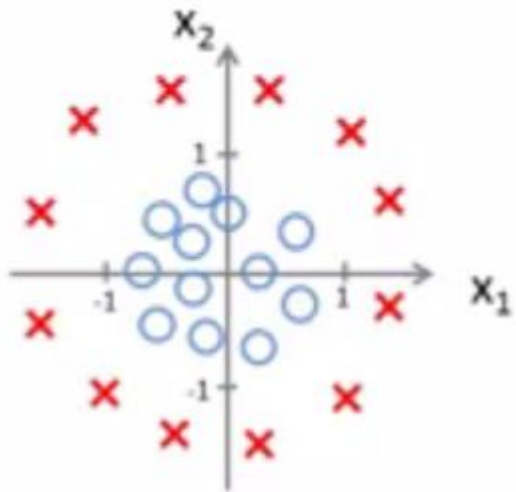


Suponha que $\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$

predizer $y = 1$ se $-3 + x_1 + x_2 \geq 0$ ou $x_1 + x_2 \geq 3$

Fronteira Não-linear

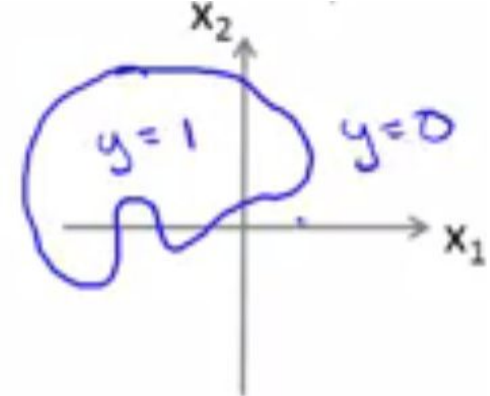
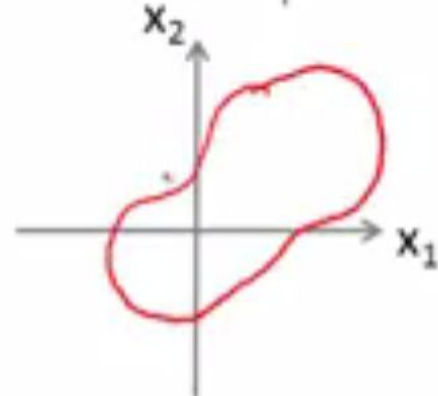
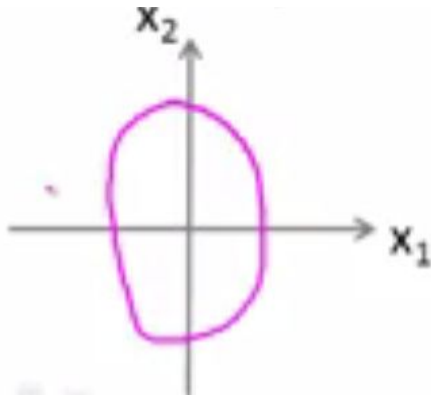
$$h_{\theta}(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$



Suponha que $\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

predizer $y = 1$ se $-1 + x_1^2 + x_2^2 \geq 0$ ou $x_1^2 + x_2^2 \geq 1$

Fronteira Não-linear





Seleção de parâmetros

- Dados:

- conjunto de treinamento $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$

- m exemplos $x \in \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_n \end{bmatrix} \quad x_0 = 1, y \in \{0, 1\}$

- forma da hipótese $h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$

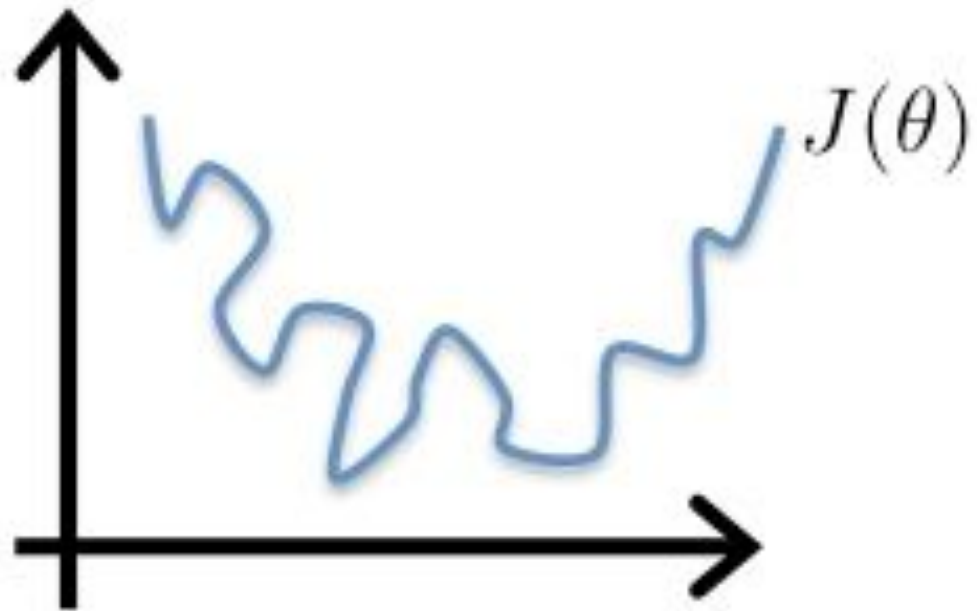
- Como selecionar o parâmetros do modelo?



Função de Custo - Regressão Linear

- A abordagem na página anterior não funciona para a regressão logística, porque:
 - A função de custo resultante seria **não linear**:
 - Não haveria garantias de o GD encontrar o mínimo global.

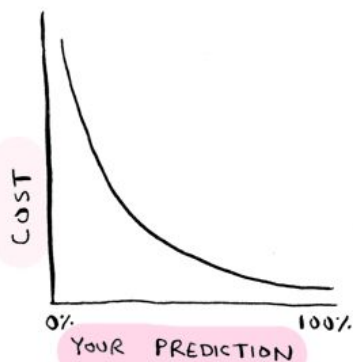
Função não-convexa



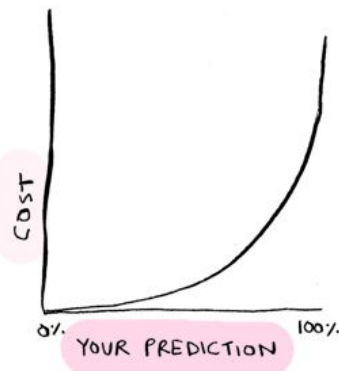
Função de Custo - Regressão Logística

$$\text{Cost}(h_{\theta}(x), y) = \begin{cases} -\log(h_{\theta}(x)) & \text{if } y = 1 \\ -\log(1 - h_{\theta}(x)) & \text{if } y = 0 \end{cases}$$

IF y is 1...



IF y is 0...





Função de custo

- Repare que o custo é zero quando a predição e a hipótese combinam:

$$(h_{\theta} = 1) \text{ e } (y = 1) \implies \text{Custo} = 0$$

$$(h_{\theta} = 0) \text{ e } (y = 0) \implies \text{Custo} = 0$$

- Entretanto, se predição e hipótese destoam:

$$(h_{\theta} \rightarrow 0) \text{ e } (y = 1) \implies \text{Custo} \rightarrow \infty$$

$$(h_{\theta} \rightarrow 1) \text{ e } (y = 0) \implies \text{Custo} \rightarrow \infty$$

- Portanto, o algoritmo é **penalizado** quando faz a predição errada.



Simplificação da função de custo

$$\text{Custo}(h_{\theta}(x), y) = -y \log(h_{\theta}(x)) - (1 - y) \log(1 - h_{\theta}(x))$$

- Repare que
quando $y = 1$ a segunda parcela é igual a zero.
quando $y = 0$ a primeira parcela é igual a zero.
- Então, podemos simplificar a função de custo:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

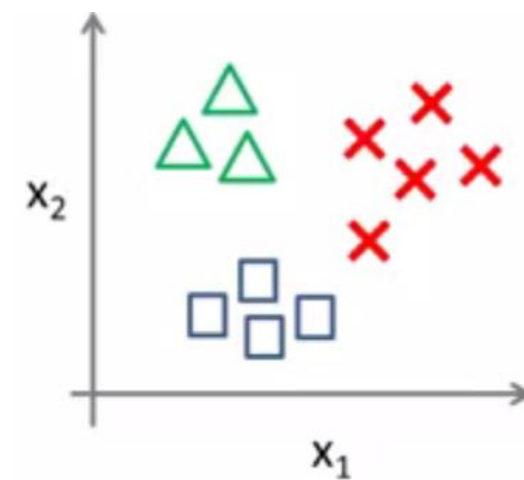
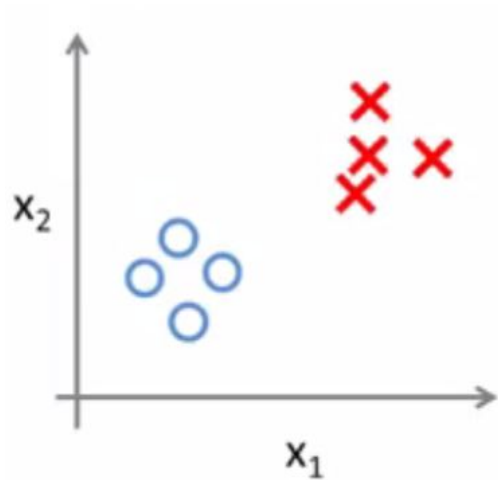




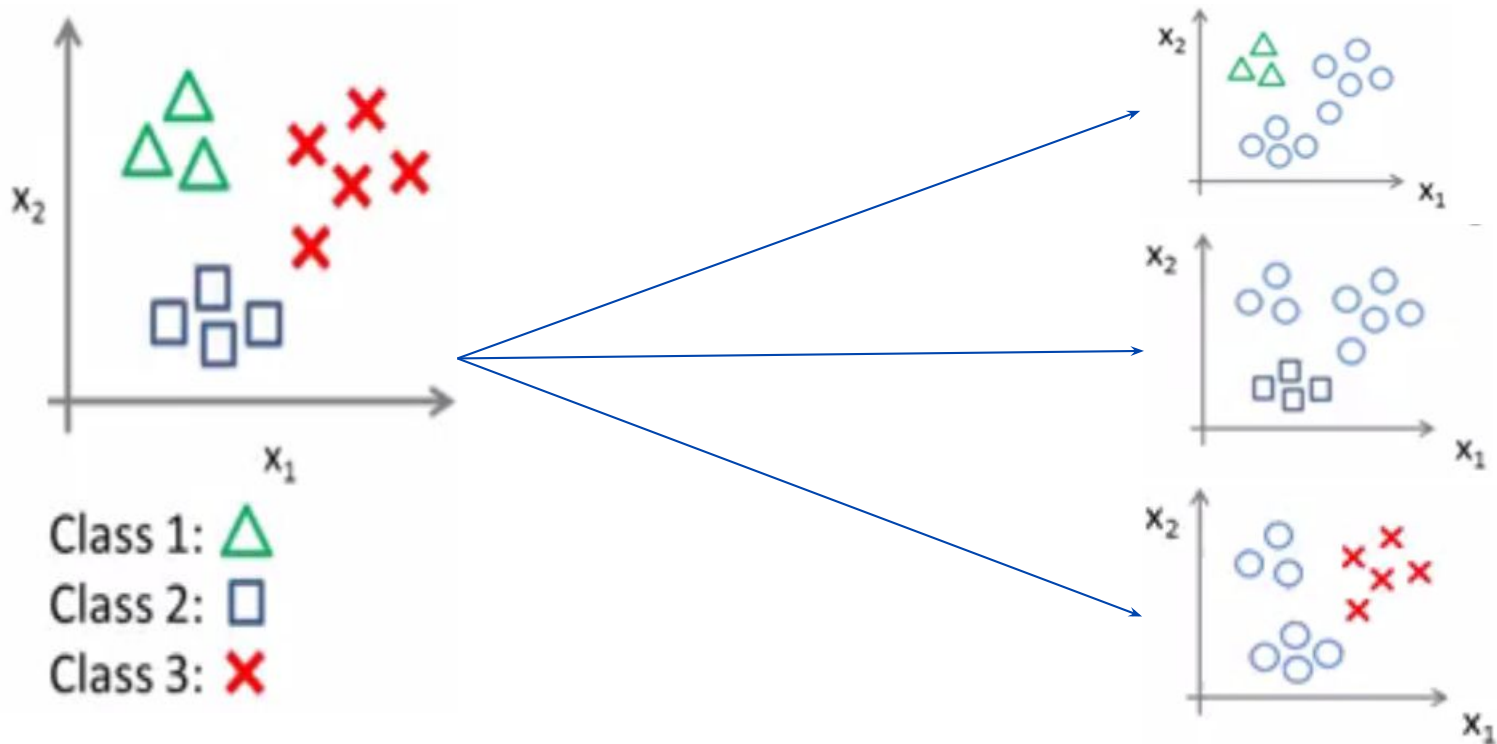
Abordagem multiclasse

- Organização de um portal de notícias: esportes, humor, política
- Diagnose médica: gripado, resfriado, dengue
- Condição do tempo: ensolarado, nublado, chuvoso

Abordagem multiclasse



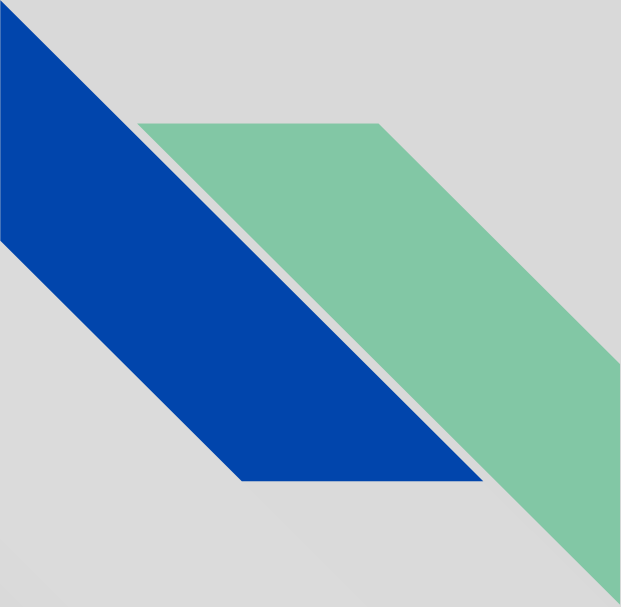
Abordagem multiclasse





Procedimento multiclasse

- Em geral para um problema de classificação com n classes, os passos são:
 - Treinar um classificador para cada uma das n classes
 - Para fazer a classificação de um novo exemplo x , seleccionar a classe que maximiza a hipótese correspondente.



Avaliação

"No futuro, o pensamento estatístico será tão necessário para a cidadania eficiente como saber ler e escrever" *H.G.Wells*



Avaliação de Modelo

- A avaliação dos classificadores binários compara dois métodos de atribuição de um atributo binário, um dos quais é geralmente um método padrão e o outro está sendo investigado.



Avaliação de Modelo

- Existem muitas métricas que podem ser usadas para medir o desempenho de um classificador ou preditor e campos diferentes têm preferências diferentes para métricas específicas devido a objetivos diferentes.



Avaliação de Modelo

- Na **medicina**, sensibilidade e especificidade são frequentemente usadas
- Na **ciência da computação**, a precisão e a recordação são preferidas.



Avaliação de Modelo

- Na **medicina**, sensibilidade e especificidade são frequentemente usadas
- Na **ciência da computação**, a precisão e a recordação são preferidas.



Definições

- **TP** = Verdadeiro Positivo (***True Positive***)
- **TN** = Verdadeiro Negativo (***True Negative***)
- **FP** = Falso Positivo (***False Positive***)
- **FN** = Falso Negativo (***False Negative***)



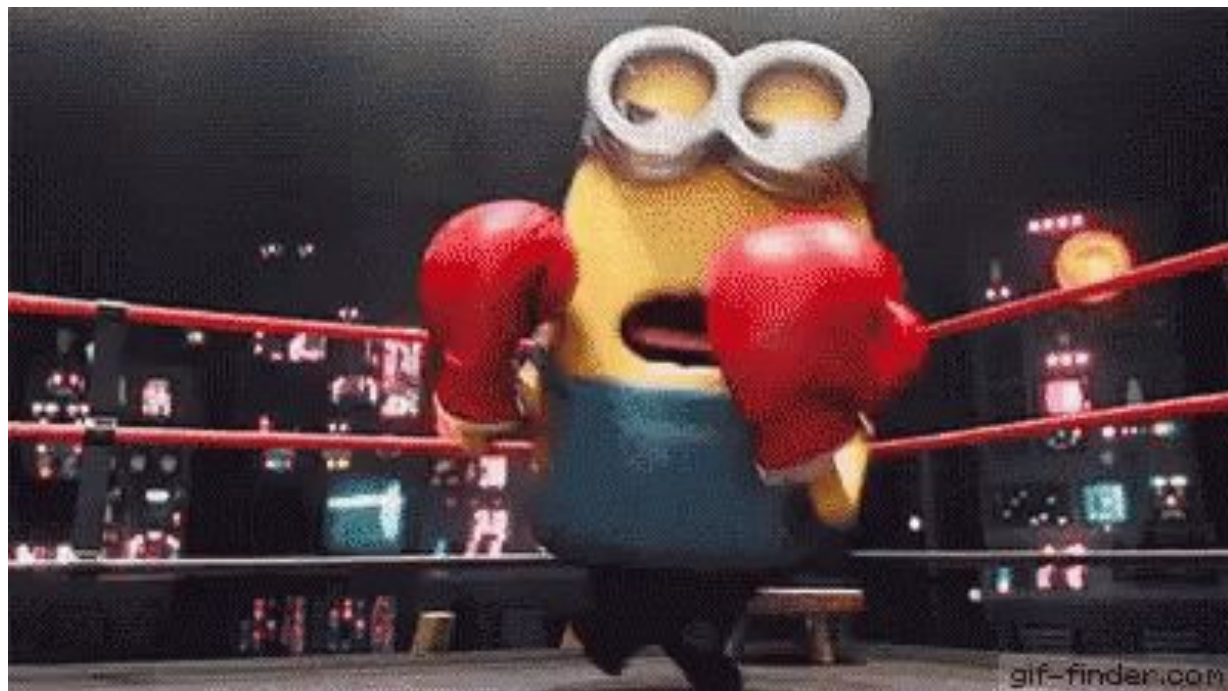
Métricas de Avaliação (Classificação)

- **Acurácia**
- **Medida F1 (*F-Score*)**
- **Matriz de Confusão**
- **Curva ROC**



Acurácia

- Segundo a Física, é a exatidão de uma medição ou de um instrumento de medição.
- Para o AM, é uma métrica para avaliar modelos de classificação.
- Pode-se dizer, informalmente, que acurácia é a fração de predições que nosso modelo acertou. Formalmente, a acurácia tem a seguinte definição:
- Acuracia = **$(TP + TN)/(TP+TN+FP+FN)$**





Especificidade

- É também chamada de taxa negativa real.
- Visa medir a proporção de negativos reais que foram corretamente classificados.
- Especificidade é uma medida estatística de desempenho para avaliar uma função de classificação. Formalmente, a especificidade tem a seguinte definição:
- Especificidade = **$TP / (TP+FP)$**



Recordação (*Recall*)

- É também conhecida como a taxa de verdadeiros positivos ou sensibilidade.
- É a soma de verdadeiros positivos e falsos negativos, que são itens que não foram rotulados como pertencentes aos positivos. Formalmente, a recordação tem a seguinte definição:
- Recordacao = **TP / (TP+FN)**



Medida F1 (*F-Score*)

- Na avaliação estatística de Classificação Binária, a medida F1 (também *F-score* ou *F-measure*) é uma medida utilizada na precisão de um teste.
- F1 tem sido amplamente utilizado na literatura de processamento de linguagem natural, como a avaliação do reconhecimento de entidade nomeada e segmentação de palavras.
- F1 é a média harmônica da **precisão** e da **recordação**, em que a pontuação de F1 atinge seu melhor valor em **1** (precisão e recordação perfeitas) e a pior em **0**. Formalmente, F1 tem a seguinte definição:
- $$F1 = 2 * (\text{Precisao} * \text{Recordacao}) / (\text{Precisao} + \text{Recordacao})$$



Coeficiente de Correlação de Matthews (MCC)

- O MCC é usado em AM como uma medida da qualidade das classificações binárias.
- Foi um método introduzido pelo bioquímico Brian W. Matthews em 1975.
- É um coeficiente de correlação entre as classificações binárias observadas e preditas, retornando um valor entre **-1** (discordância total) e **+1** (previsão perfeita). Formalmente, a acurácia tem a seguinte definição:
- $$\text{MCC} = \frac{(\text{TP} \cdot \text{TN}) - (\text{FP} \cdot \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$



Matriz de Confusão

- Para problemas de classificação em AM, a matriz de confusão é também conhecida como **matriz de erro**, com um *layout* de tabela específico que permite a visualização do desempenho de um algoritmo, tipicamente em Aprendizado Supervisionado.
- Em Aprendizado Não-Supervisionado, é geralmente chamada de **matriz de correspondência**.
- É uma tabela com duas dimensões e conjuntos classes em ambas as dimensões.



Matriz de Confusão

		Valor Previsto	
		Positivo	Negativo
Valor Verdadeiro	Negativo	Verdadeiros Positivos	Falsos Negativos
	Positivo	Falsos Positivos	Verdadeiros Negativos

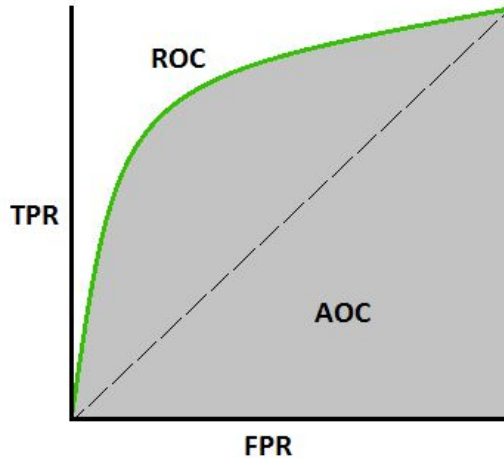


Curva ROC

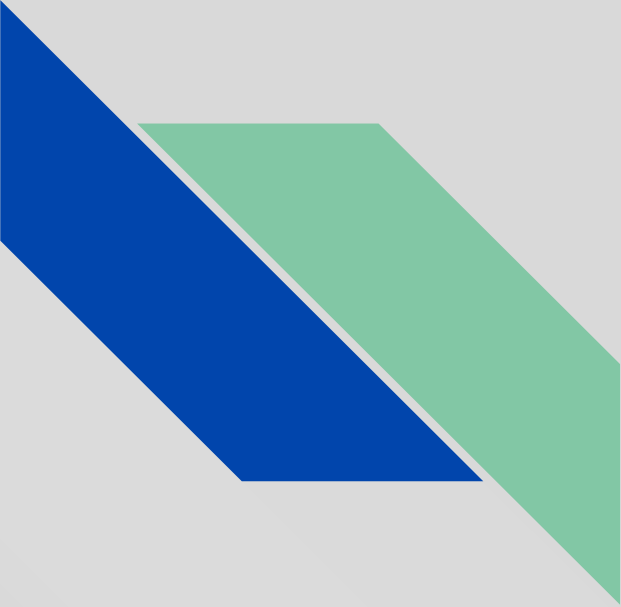
- É uma representação gráfica que ilustra o desempenho de um classificador binário e como o seu limiar de discriminação é variado.
- A curva ROC foi desenvolvida por engenheiros durante a Segunda Guerra Mundial para detecção de objetos inimigos nas batalhas e já foi implementado na psicologia, medicina, radiologia, aprendizado de máquina e mineração de dados.
- Esta métrica está relacionada com a análise de custo/benefício do diagnóstico da Tomada de Decisão.

Curva ROC

- É obtido pela representação da fração de Positivos Verdadeiros dos Positivos Totais ($TPR = PV/P$) versus a fração de Positivos Falsos dos Negativos Totais ($FPR = PF/N$), em várias configurações do limite. Formalmente, a Curva ROC pode ser expressa da seguinte forma:







Técnicas

"Não são tanto as coisas que não sabemos que nos metem em confusões. São as coisas que pensamos que sabemos." *Artemus Ward*



Aprendizado de Máquina Supervisionado

- Aprendizado de máquina (AM) é o estudo científico de algoritmos e modelos estatísticos que os sistemas de computador usam para realizar uma tarefa específica sem usar instruções explícitas, confiando em padrões e inferência



Técnicas de Classificação

- Técnicas de classificação constróem um modelo matemático com base nos dados de amostra e seus rótulos, para fazer previsões ou conduzir decisões sem ser explicitamente programado para executar as tarefas



Técnicas de Classificação

- Naive Bayes
- Árvore de Decisão
- Regressão Logística
- Máquina de Vetor de Suporte
- Florestas Aleatórias
- Aprendizado baseado em instância



Naive Bayes

Por Gustavo Alexandre



Naive Bayes

O **Naive Bayes** é um algoritmo probabilístico simples baseado no teorema de Bayes.

Este utiliza dados de treino para formar um modelo probabilístico baseado na evidência das features nos dados.

O algoritmo supõe que há uma independência entre as features do modelo.

Isso significa que a presença de uma determinada feature não tem nenhuma relação com as outras.



Casos de Uso

O **Naive Bayes** é um algoritmo probabilístico simples baseado no teorema de Bayes.

Este utiliza dados de treino para formar um modelo probabilístico baseado na evidência das features nos dados.

O algoritmo supõe que há uma independência entre as features do modelo.

Isso significa que a presença de uma determinada feature não tem nenhuma relação com as outras.





Árvore de Decisão

Por Gustavo Alexandre



Árvore de Decisão

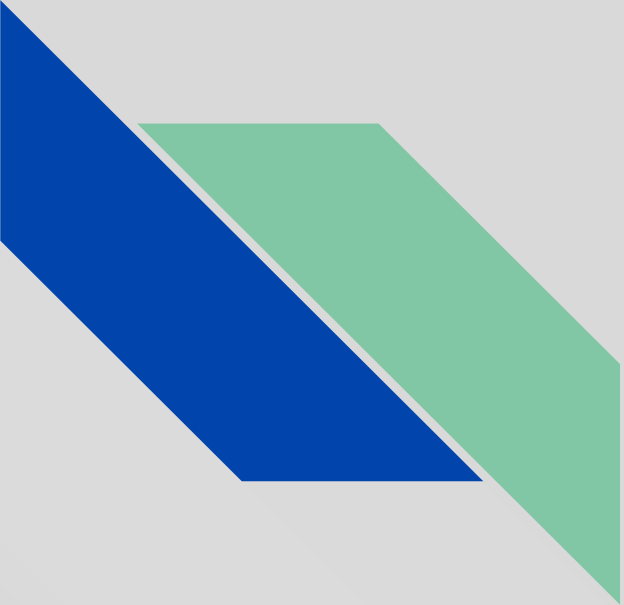
Uma árvore de decisão é uma ferramenta de **suporte à tomada de decisão** que usa um gráfico no formato de árvore e demonstra visualmente as condições e as probabilidades para se chegar a resultados. O algoritmo utilizado para chegar na representação visual da árvore pertence ao grupo de **aprendizado de máquina supervisionado**, e funciona tanto para **regressão** quanto para **classificação**.



Árvore de Decisão

Um árvore de decisão é uma forma de **visualizar as regras de negócio** que levam a determinados grupos de indivíduos, construídos com base em uma variável *target*. Por exemplo: “Quais são as regras, baseadas em variáveis pré-especificadas no modelo, que levam certos grupos de clientes a apresentarem uma menor taxa de *churn*, e outros uma taxa muito mais elevada?”





Regressão Logística

(Logit)

Por Gustavo Alexandre



Regressão Logística

- **Regressão logística** é uma técnica estatística muito poderosa, utilizada para modelagem de saídas binárias (sim ou não). Quando se quer medir a relação de uma variável dependente binária com uma ou mais variáveis independentes, é comum utilizar esta técnica.
- Pense, por exemplo, numa empresa que empresta dinheiro para um cliente. Com base nas informações deste cliente (idade, profissão, etc.), é interessante a empresa tentar prever se o cliente vai pagar a dívida ou não. Uma forma de tentar prever isso é utilizando a regressão logística.



Regressão Logística

- A **regressão logística** é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias^{[1][2]}
- A regressão logística é amplamente usada em ciências médicas e sociais, e tem outras denominações, como **modelo logístico**, **modelo logit**, e **classificador de máxima entropia**.



Regressão Logística

- Em comparação com as técnicas conhecidas em regressão, em especial a regressão linear, a regressão logística distingue-se essencialmente pelo facto de a variável resposta ser categórica.
- Enquanto método de predição para variáveis categóricas, a regressão logística é comparável às técnicas supervisionadas propostas em aprendizagem automática (árvores de decisão, redes neurais, etc.), ou ainda a análise discriminante preditiva em estatística exploratória. É possível de as colocar em concorrência para escolha do modelo mais adaptado para um certo problema preditivo a resolver.
- Trata-se de um modelo de regressão para variáveis dependentes ou de resposta binomialmente distribuídas. É útil para modelar a probabilidade de um evento ocorrer como função de outros factores. É um modelo linear generalizado que usa como função de ligação a função logit.

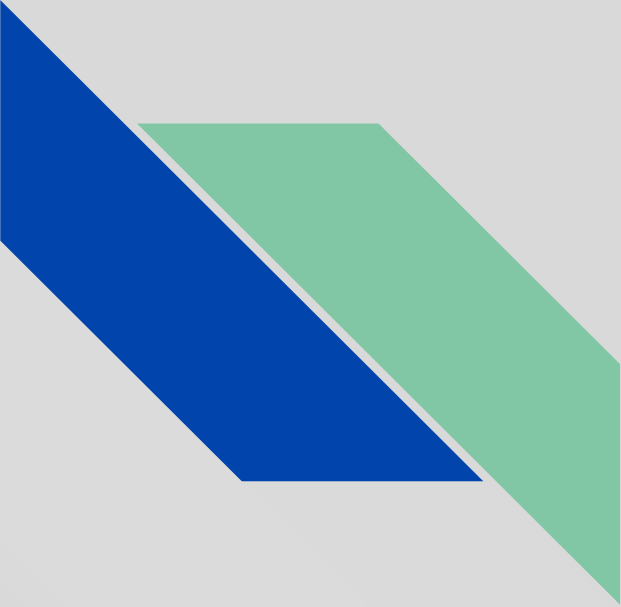


Casos de Uso

A regressão logística é utilizada em áreas como as seguintes:

- Em medicina, permite por exemplo determinar os factores que caracterizam um grupo de indivíduos doentes em relação a indivíduos sãos.
- No domínio dos seguros, permite encontrar frações da clientela que sejam sensíveis a determinada política securitária em relação a um dado risco particular.
- Em instituições financeiras, pode detectar os grupos de risco para a subscrição de um crédito.
- Em econometria, permite explicar uma variável discreta, como por exemplo as intenções de voto em actos eleitorais.





Máquina de Vetor de Suporte (SVM)



Máquina de Vetor de Suporte

- Uma **máquina de vetores de suporte** (SVM, do inglês *support vector machine*) é um conceito na ciência da computação para um conjunto de métodos do aprendizado supervisionado que analisam os dados e reconhecem padrões, usado para classificação e análise de regressão.
- O SVM padrão toma como entrada um conjunto de dados e prediz, para cada uma entrada, qual das possíveis classes essa entrada faz parte, o que faz do SVM um classificador linear binário não probabilístico.
- Dado um conjunto de exemplos de treinamento, cada um marcado como pertencente a uma de duas categorias, um algoritmo de treinamento do SVM constrói um modelo que atribui novos exemplos a uma categoria ou outra.

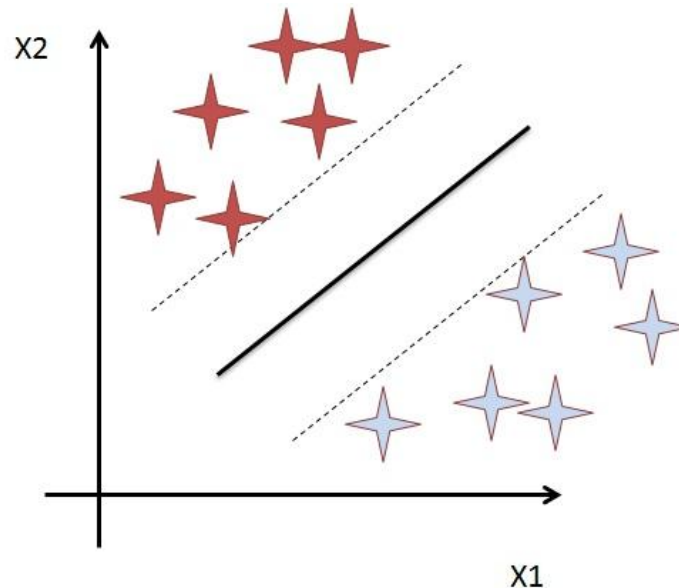


Máquina de Vetor de Suporte

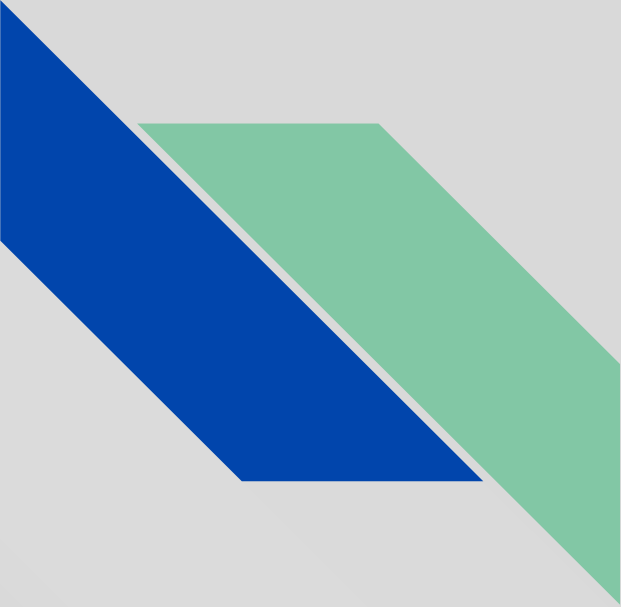
- Um modelo SVM é uma representação de exemplos como pontos no espaço, mapeados de maneira que os exemplos de cada categoria sejam divididos por um espaço claro que seja tão amplo quanto possível.
- Os novos exemplos são então mapeados no mesmo espaço e preditos como pertencentes a uma categoria baseados em qual o lado do espaço eles são colocados.

Máquina de Vetor de Suporte

- SVM busca um *hiperplano* entre os dados a serem classificados e visa maximizar a distância entre os pontos, separando cada uma das classes:







Seleção de Atributos

(Feature Selection)



Seleção de Atributos

No aprendizado de máquina e estatística, a seleção de recursos, também conhecida como seleção de variáveis, seleção de atributos ou seleção de subconjuntos de variáveis, é o processo de seleção de um subconjunto de recursos relevantes (variáveis e preditores) para uso na construção de modelos. As técnicas de seleção de recursos são usadas por quatro razões:

- simplificação de modelos para torná-los mais fáceis de interpretar por pesquisadores/usuários
- menor tempo de treinamento
- evita a alta dimensionalidade
- generalização melhorada reduzindo *overfitting* (formalmente, redução de variância)



Seleção de Atributos

- A premissa central ao aplicar **seleção de atributos** é que os dados podem conter atributos redundantes ou irrelevantes e, portanto, são possíveis de remoção sem incorrer perdas significativas de informações.
- Redundantes e irrelevantes são duas noções distintas, uma vez que uma característica relevante pode ser redundante na presença de outra característica relevante com a qual ela é fortemente correlacionada.



Seleção de Atributos

- As técnicas de **seleção de atributos** devem ser diferenciadas da extração de recursos.
- A extração de recursos cria novos recursos a partir das funções dos recursos originais, enquanto a seleção de recursos retorna um subconjunto dos recursos.
- As técnicas de **seleção de atributos** são frequentemente usadas em domínios em que há muitos recursos e comparativamente poucas amostras.
- Casos arquetípicos para a aplicação de **seleção de atributos** incluem a análise de textos escritos e dados de DNA, onde existem milhares de atributos, e algumas dezenas de centenas de amostras.





Florestas Aleatórias

(Random Forests)



Florestas Aleatórias

- Florestas aleatórias (*Random Forest*) são um método de aprendizado conjunto para classificação, regressão e outras tarefas que operam construindo uma multiplicidade de árvores de decisão no momento do treinamento e gerando a classe que é o modo das classes (classificação) ou previsão média (regressão) das árvores individuais.
- Florestas de decisão aleatória corrigem o hábito de **overfitting** de árvores de decisão em seu conjunto de treinamento
- O primeiro algoritmo para florestas de decisão aleatória foi criado por [Tin Kam Ho](#) usando o método de subespaço aleatório, que, na formulação de Ho, é uma maneira de implementar a abordagem de discriminação estocástica para classificação proposta por Eugene Kleinberg.



Florestas Aleatórias

- Uma extensão do algoritmo foi desenvolvida por Leo Breiman e Adele Cutler, que registrou "[Random Forests](#)" como uma marca comercial (em 2019, de propriedade da Minitab, Inc.).
- A extensão combinou a idéia de "ensacamento" de Breiman e a seleção aleatória de características, introduzida primeiramente por Ho e, posteriormente, Amit e Geman, a fim de construir uma coleção de árvores de decisão com variância controlada [[breiman2001](#)].



Florestas Aleatórias

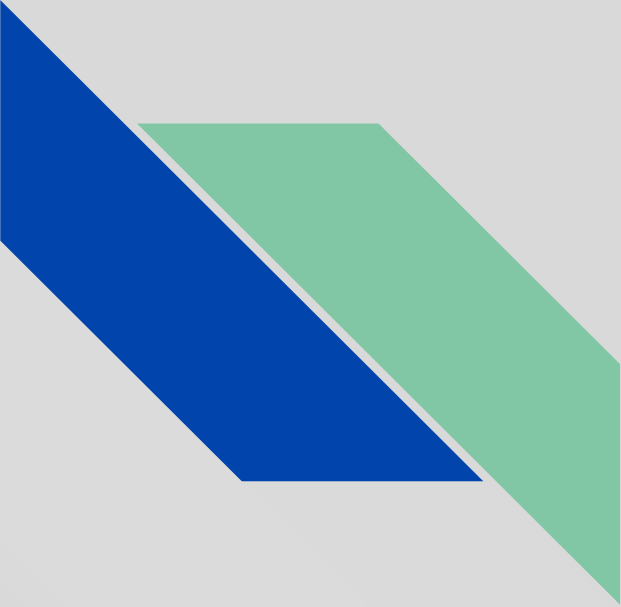
- Essa observação de um classificador mais complexo (uma floresta maior) vai ficando mais precisa quase monotonicamente está em nítido contraste com a crença comum de que a complexidade de um classificador só pode crescer até um certo nível de precisão antes de ser ferida pelo *overfitting*.
- A explicação da resistência do método *Random Forest* ao overtraining pode ser encontrada na teoria da discriminação estocástica de Kleinberg [[kleinberg1996](#)].



Florestas Aleatórias

- A introdução de florestas aleatórias propriamente ditas foi feita primeiramente em um artigo de Leo Breiman, através de um método de construção de uma floresta de árvores não correlacionadas usando um procedimento semelhante ao CART, combinado com otimização de nó aleatório e ensacamento. Além disso, este trabalho combina alguns novos, que formam a base da prática moderna de florestas aleatórias, em particular:
 - Usando erro *out-of-bag* como uma estimativa do erro de generalização.
 - Medindo a importância variável através de permutação.
- O relatório também oferece o primeiro resultado teórico para florestas aleatórias na forma de um limite no erro de generalização que depende da força das árvores na floresta e sua correlação.





Aprendizado baseado em Instância

Por Gustavo Alexandre



Aprendizagem Baseada em Instâncias

- No aprendizado de máquina, a aprendizagem baseada em instância é uma família de algoritmos de aprendizado que compara novas instâncias de problemas com instâncias vistas em treinamento, que foram armazenadas na memória.
- É chamado de baseado em instância porque constrói hipóteses diretamente das próprias instâncias de treinamento. Isso significa que a complexidade da hipótese pode crescer com os dados; no pior caso, uma hipótese é uma lista de n itens de treinamento e a complexidade computacional de classificar uma única nova instância é $O(n)$.



Aprendizagem Baseada em Instâncias

- Uma vantagem que o aprendizado baseado em instância tem sobre outros métodos de aprendizado de máquina é sua capacidade de adaptar seu modelo a dados nunca vistos antes.
- Ou seja, pode simplesmente armazenar uma nova instância ou descartar uma instância antiga.
- Exemplos de algoritmo de aprendizado baseado em instâncias são o algoritmo KNN, máquinas kernel e redes RBF.



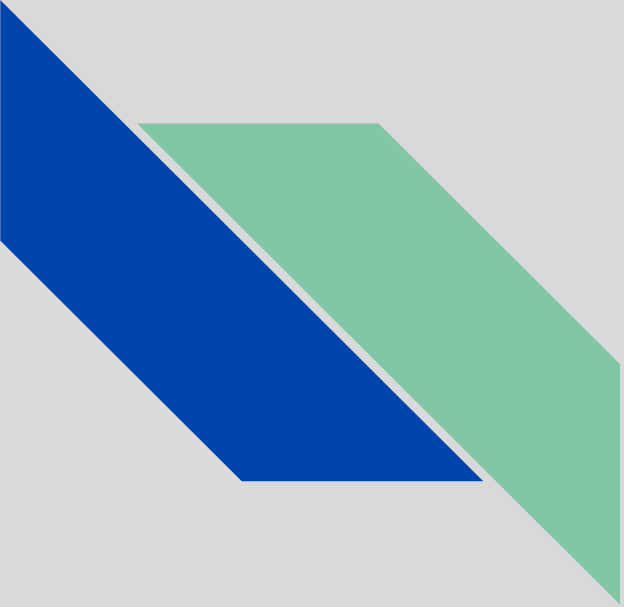
Aprendizagem Baseada em Instâncias

- Os algoritmos (KNN, RBF e máquinas de Kernel) armazenam seu conjunto de treinamento ao prever um valor/classe para uma nova instância.
- Esses algoritmos calculam distâncias ou semelhanças, dada uma instância a ser avaliada e as instâncias de treinamento já identificadas.
- Para combater a complexidade da memória de armazenar todas as instâncias de treinamento, bem como o risco de *overfitting* ao ruído no conjunto de treinamento, algoritmos de redução de instâncias foram propostos.



Casos de Uso

- Gagliardi aplica essa família de classificadores na área médica como ferramentas de diagnóstico de segunda opinião e como ferramentas para a fase de extração de conhecimento no processo de descoberta de conhecimento em bancos de dados.
- Um desses classificadores (denominado classificador de aprendizado exemplar Prototype (PEL-C) é capaz de extrair uma mistura de casos prototípicos abstratos (que são síndromes) e casos clínicos atípicos selecionados.



Considerações

"Alguns usam a estatística como os bêbados usam postes: mais para apoio do que para iluminação" *Andrew Lang*

