



O que é inteligência artificial?

O que é inteligência artificial?

O conceito de inteligência artificial



Área da computação que se dedica a estudar e propor mecanismos que façam o computador simular uma habilidade natural presente nos seres humanos de pensar de forma inteligente e autônoma, raciocinando;

As técnicas de inteligência artificial são propostas para trabalhar de forma independente diante de problemas que são apresentados, devendo aprender e construir o conhecimento necessário para resolvê-los.

O que é inteligência artificial?

O conceito de inteligência artificial



O desenvolvimento da inteligência artificial se expandiu junto com as tecnologias computacionais que são responsáveis por viabilizar novas aplicações e o processamento de grandes quantidades de dados;

Com isso, os recursos da IA foram se tornando cada vez mais abrangentes e hoje são utilizados em muitas áreas: jogos, aplicações empresariais, robótica, sistemas de segurança, saúde, entre outras.

O que é inteligência artificial?

O conceito de inteligência artificial



O objetivo de criar sistemas computacionais inteligentes é que os computadores tem uma capacidade muito mais elevada e precisa para analisar grandes quantidades de dados, assim eles podem tomar e apoiar decisões embasadas.

Percebendo o potencial da inteligência artificial começou-se a vislumbrar outras utilidades para ela, partindo para simulação de outras capacidades humanas de dialogar, ter sentimentos, etc.

Os chatbots são grandes exemplos da inteligência artificial buscando se desenvolver no aspecto do processamento de linguagem natural para criar sistemas autônomos para conversar com pessoas com uma linguagem próxima da humana.

O que é inteligência artificial?

A inteligência artificial e indústria 4.0



Os benefícios da inteligência artificial são muitos, especialmente para aqueles que possuem recursos computacionais avançados e podem explorar dos seus préstimos em diferente frentes.

Com muitos recursos e sempre buscando se manter atualizadas no mercado que atuam, as indústrias estão na vanguarda no uso da inteligência artificial e outras tecnologias, formando o fenômeno chamado de indústria 4.0.

As transformações pelas tecnologias dentro dos ambientes industriais são tão grandes que alguns pesquisadores defendem que estamos vivendo uma nova revolução industrial, a quarta.

O que é inteligência artificial?

Técnicas da inteligência artificial



Como exemplos de técnicas de inteligência artificial temos:

- redes neurais;
- algoritmos genéticos;
- árvores de decisão;
- regras de associação.

Muita confusão é causada pelo fato dessas técnicas também serem apontadas para a mineração de dados e aprendizado de máquina. Isso é correto porque essas técnicas são projetadas para simular o raciocínio, conforme critérios da IA e são utilizadas para aprender e explorar dados, como previsto no Machine Learning e Mineração de dados.

O que é inteligência artificial?

Presença da inteligência artificial

A inteligência artificial está presente em muitas coisas no nosso dia a dia:

- nos sistemas de recomendação de músicas e filmes;
- dentro dos buscadores web;
- no funcionamento das redes sociais;
- entre outros.

O que é inteligência artificial?

As controvérsias da IA



Para muitas pessoas existe um receio com relação a inteligência artificial, com medo que ela venha a prejudicar a existência de vagas de trabalho porque irá substituir o ser humano em muitas funções.

Esse é um termo justificável para aqueles que realizam tarefas muito mecanizadas, porque a IA com outras técnicas, como a automação, poderá substituir esse tipo de mão de obra. Por outro lado, novos postos de trabalho serão criados requerendo uma mão de obra mais técnica para atuar no desenvolvimento de novas tecnologias e soluções.

O que é inteligência artificial?

As controvérsias da IA



Uma outra questão muito importante com relação a IA é a possibilidade dela se desenvolver tanto ao ponto de colocar em risco a vida humana.

Esse pensamento surge por conta da possibilidade de se criar máquinas inteligentes e autônomas, que também contam com grande poder físico. Pelo fato delas serem autônomas teme-se que elas assumam o controle de tudo que existe e sinta que os homens representam uma ameaça para a soberania delas.



Obrigada!

Ana Laurentino



Aprendizado de máquina

O conceito de aprendizado de máquina

O aprendizado de máquina ou *machine learning* é um termo que está ficando cada vez mais conhecido, assim como, a inteligência artificial.

Esse processo faz todo sentido, uma vez que, o aprendizado de máquina é uma subárea da inteligência artificial.

Isso fica evidente quando pensamos que o objetivo do aprendizado de máquina é fazer máquinas aprenderem, mas para aprender é preciso raciocinar que o objetivo da inteligência artificial.

Portanto, qualquer abordagem usada para aprendizagem de máquina é também uma abordagem da inteligência artificial.

O conceito de aprendizado de máquina

O aprendizado de máquina trabalha com a ideia de que é possível fazer com que máquinas simulem o processo de aprendizado humano.

As formas de aprender podem acontecer de diferentes maneiras:

- com base em experiências passadas;
- realizando comparações;
- inferindo ou identificando relações.

De uma forma mais objetiva, os computadores podem utilizar dados para realizar diferentes análises e através delas aprender.

O objetivo do aprendizado de máquina

O aprendizado de máquina realiza muitas análises nos dados e internaliza os resultados obtidos para realizar previsões e apoiar decisões.

É possível perceber que as tecnologias que trabalham com dados, realizando análises, gerando informações e conhecimentos, são todas projetadas e utilizadas para dar suporte a tomada de decisões.

Para qualquer negócio, tomar boas decisões é fundamental para sucesso, o que fica muito mais quando sustentado por boas informações.

Categorias do aprendizado de máquina

O aprendizado de máquina divide as suas técnicas em duas categorias básicas, segundo a forma como os algoritmos aprendem:

- aprendizado supervisionado;
- aprendizado não supervisionado.

Atualmente é possível encontrar outras variações de aprendizado, que misturam características dos dois, mas são esses os principais.

O aprendizado supervisionado

O aprendizado supervisionado é feito quando além de possuir um conjunto de dados para alimentar o algoritmo, existe um conjunto de dados com as saídas esperadas.

Assim, a solução que estiver sendo criada pode ser trabalhada em duas etapas:

- treinamento;
- teste.

O aprendizado supervisionado - Treinamento

Na etapa de treinamento o algoritmo utiliza os dados de entrada para realizar suas análises e depois os dados de saída para verificar se as análises que realizou estão corretas.

Contando com dados para treinamento, o algoritmo pode saberá se está acertando ou errando muito, com base nisso pode ser aprimorado até que se atinja um nível desejado de qualidade para se tornar uma solução que possa ser colocada em prática.

O aprendizado supervisionado - Teste

Depois de terminado o treinamento, inicia-se a etapa de testes onde o algoritmo é colocado a prova analisando dados de entrada desconhecidos e sem o oferecimento da saída esperada ou a possibilidade de se aprimorar.

Nesse momento, deve ser realizada uma validação nos resultados que o algoritmo apresentou, para que saiba se o modelo está adequado para ser colocado em prática.

O aprendizado supervisionado - Teste



Bonin (2018)

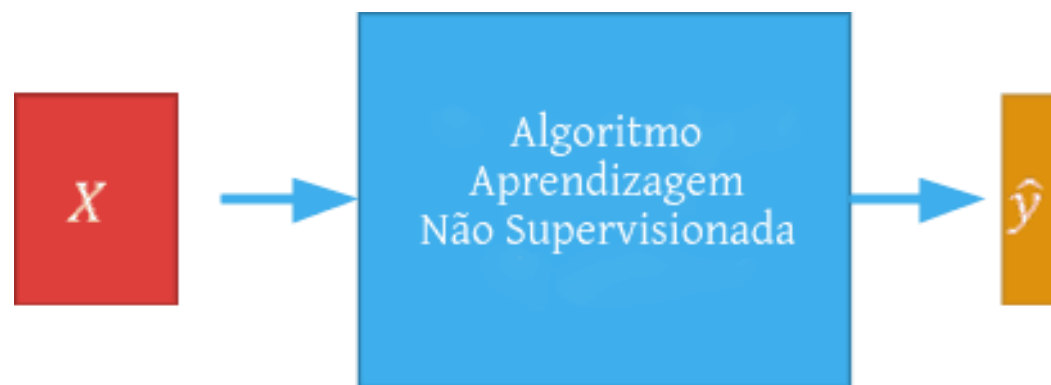
O aprendizado não supervisionado

Quando falamos do aprendizado não supervisionado estamos dizendo que o algoritmo deverá aprender com os dados, mas sem contar com as saídas para validar o trabalho que está executando.

Sem os dados os algoritmos precisam aprender apenas com base nas inferências, relações e padrões de comportamento que podem ser percebidos no conjunto de dados de entrada.

No aprendizado não supervisionado não existe uma etapa específica para o treinamento e outra para teste.

O aprendizado não supervisionado



Bonin (2018)

Onde está o aprendizado de máquina?

O aprendizado de máquina está bastante presente no cotidiano da grande maioria das pessoas:

- nas campanhas de marketing cada vez mais específica e direcionada dentro da internet;
- nos sistemas utilizados por empresas financeiras para detectar fraudes;
- nas aplicações de comércio eletrônico para detectar fraudes;
- na otimização do estoque, entendendo o comportamento de consumo dos clientes;
- outras.

Algoritmos de aprendizado de máquina

Muitos algoritmos estão disponíveis para serem utilizados para a mineração de dados, devendo a escolha ser realizada analisando o problema e o funcionamento do algoritmo:

- árvores de decisão;
- classificador de Naive Bayes;
- método dos mínimos quadrados;
- regressão logística;
- suporte de máquinas de vetores;
- decomposição em valores singulares;
- redes neurais.

Referências

BONIN, M. Machine Learning: Uma Introdução. 2018. Disponível em:
<<https://king.host/blog/2018/03/machine-learning-uma-introducao/>>. Acesso em: 20 Jan. 2019.



Obrigada!

Ana Laurentino



O algoritmo Naive Bayes

O algoritmo Naive Bayes

O algoritmo Naive Bayes

O algoritmo chamado de Naive Bayes é um classificador construído com base no Teorema proposto por Thomas Bayes.

Uma das principais características deste algoritmo é que ele desconsidera a relação entre os diferentes atributos de um padrão de entrada, porque trabalha cada padrão de forma individual.

O funcionamento do Naive Bayes

Suponhamos que estamos trabalhando com um classificador para diagnosticar uma doença nova e para isso foram utilizados dados de 100 pessoas.

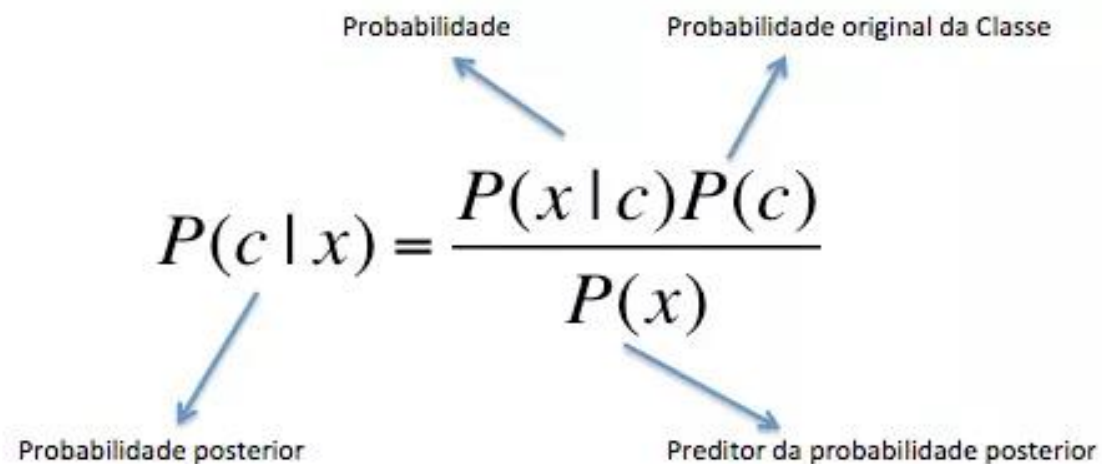
Dentre as pessoas que fizeram parte do estudo, 20% delas realmente possuía a doença e o restante, 80%, não apresentava a doença.

Em meio as pessoas que apresentavam a doença, 90% tiveram o seu teste com resultado positivo.

Aconteceu também que 30% das pessoas que não tem a doença receberam um teste sinalizando positivo.

O algoritmo Naive Bayes

A fórmula Naive Bayes



The diagram shows the Naive Bayes formula with four blue arrows pointing from labels to parts of the formula:

- Probabilidade** points to $P(x|c)$.
- Probabilidade original da Classe** points to $P(c)$.
- Probabilidade posterior** points to $P(c|x)$.
- Preditor da probabilidade posterior** points to $P(x)$.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Como responder?

Se fosse feito o seguinte questionamento:

- “Uma pessoa recebeu um teste positivo, qual a probabilidade de ela ter a doença?”

O algoritmo de Naive Bayes buscará determinar uma probabilidade a posteriori, através da multiplicação da probabilidade a priori pela de “receber um resultado positivo quando se tem a doença”.

Como responder?

Vamos aos cálculos:

- $P(\text{doença} | \text{positivo}) = 20\% * 90\%;$
- $P(\text{doença} | \text{positivo}) = 0,2 * 0,9;$
- $P(\text{doença} | \text{positivo}) = 0,18;$
- $P(\text{não doença} | \text{positivo}) = 80\% * 30\%;$
- $P(\text{não doença} | \text{positivo}) = 0,8 * 0,3;$
- $P(\text{não doença} | \text{positivo}) = 0,24.$

Como responder?

Normalizando os dados, fazendo com que as probabilidades somadas fiquem igual a 1. Para isso, o resultado é dividido pela soma das duas probabilidades:

- $P(\text{doença} | \text{positivo}) = 0,18 / (0,18 + 0,24) = 0,4285$;
- $P(\text{não doença} | \text{positivo}) = 0,24 / (0,18 + 0,24) = 0,5714$;
- $0,4285 + 0,5714 = 0,9999..$ ou aproximadamente 1.

O algoritmo Naive Bayes

Como responder?

Aplicando a fórmula apresentada, chega-se a uma probabilidade de cerca de 43% de uma pessoa ter a doença quando recebe um exame positivo.



Obrigada!

Ana Laurentino



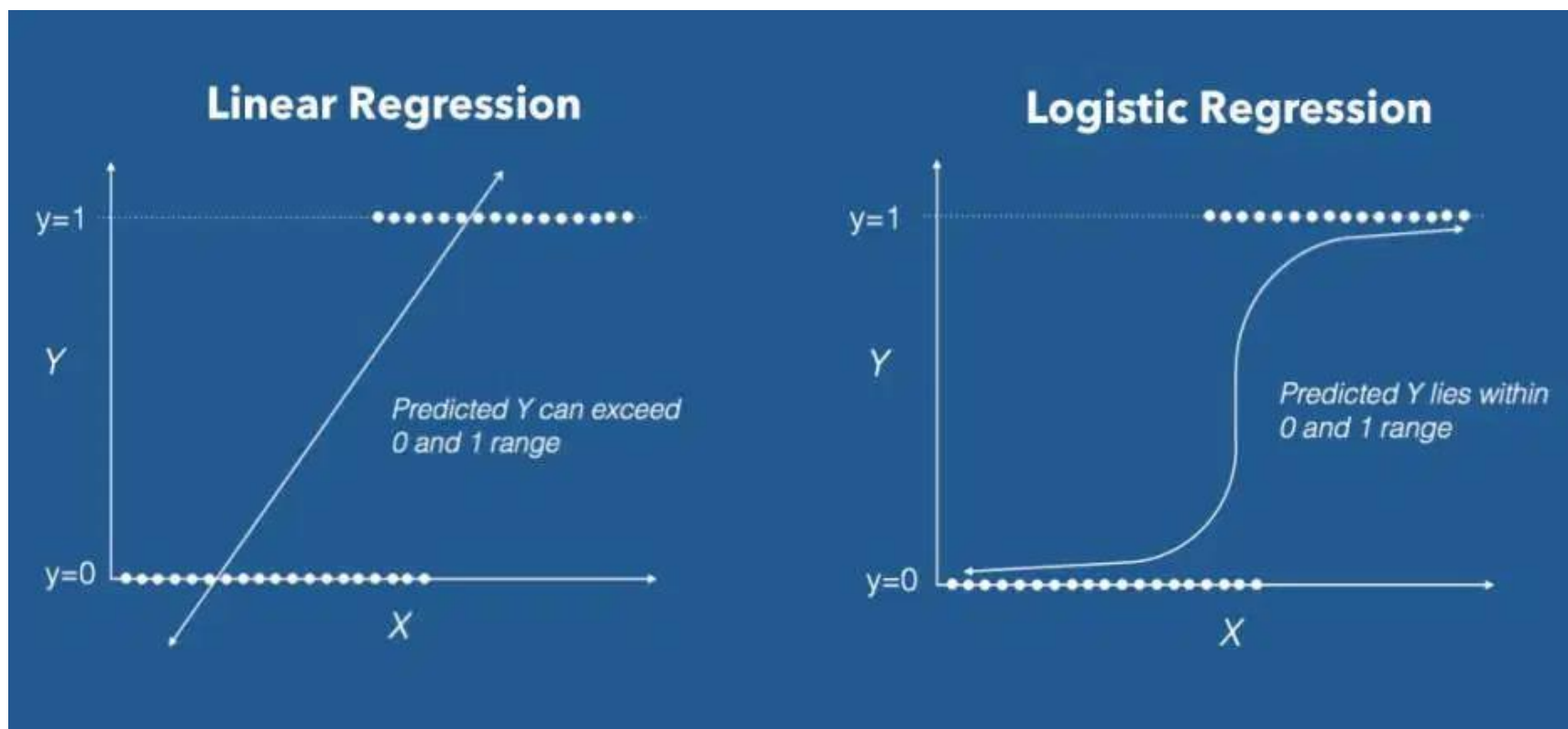
Regressão logística

A regressão logística

O algoritmo de regressão logística pode ser entendido como análogo ao de regressão linear, mas com a diferença de que pode trabalhar com uma saída expressa em valores numéricos, ao invés de apenas classes.

A regressão logística mistura conhecimentos da probabilidade e da estatística. A relação entre atributos e uma variável é dada através de uma função logística.

A regressão logística



Os tipos de regressão logística

- **Regressão logística binomial:** Os objetos são classificados em dois grupos ou categorias. É quase um processo de determinar o que é e o que não é. Se o e-mail é spam ou não, se a célula é cancerígena ou não.
- **Regressão logística ordinal:** É um modelo diferente porque trabalha com categorias ordenadas. Com esse modelo podem ser classificadas três ou mais classes que possuem uma ordem conhecida. Por exemplo, o aluno foi ruim, regular ou ótimo.
- **Regressão logística multinomial:** Neste modelo, os objetos são classificados em três ou mais categorias que não possuem ordem entre si. Por exemplo, esse alimento é uma verdura, fruta ou legume.

Onde está a regressão logística?

Um dos principais exemplos do uso da regressão logística é na filtragem de e-mails para identificação daqueles que são mensagens de spam e os que são mensagens desejadas.

Essa é uma ferramenta de segurança importante para proteção de empresas e pessoas de acessarem conteúdos que não são desejadas e que podem conter vírus.

Regressão logística

A regressão logística

<https://estatsite.com/2018/08/29/regressao-logistica-conceitos-e-formula>



Obrigada!

Ana Laurentino

Unyleya
EDUCACIONAL



KNN

O algoritmo KNN / K - Nearest Neighbor

- Uma abordagem simples para classificação de dados;
- Trabalha com o conceito de vizinhança para classificação dos dados;
- O algoritmo pode ser utilizado para regressão ou classificação com devidas adaptações.

O algoritmo KNN / K - Nearest Neighbor

No algoritmo K-NN o valor de K é a indicação de quantos vizinhos serão utilizados na definição da classe para a amostra.

Não existe uma definição sobre qual é o melhor valor para se utilizar como o K. Caso o K seja muito pequeno a classificação estará muito sensível a ruídos, se o K for muito grande, poderá selecionar representantes de todas as classes.

O algoritmo KNN / K - Nearest Neighbor

O algoritmo k-NN consiste em, para cada nova amostra:

1. Calcular a distância entre o exemplo desconhecido e os outros do conjunto de treinamento;
2. Identificar os K vizinhos mais próximos;
3. A amostra será rotulada com a classe mais representada por seus vizinhos.

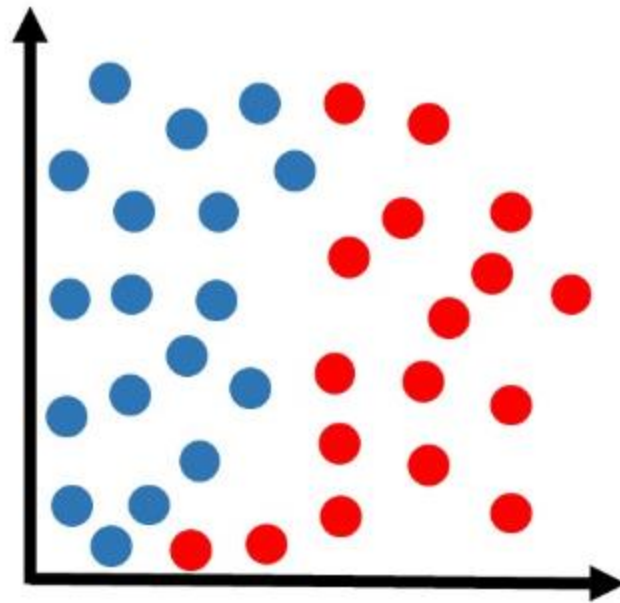
Métricas

Para o cálculo da distância podem ser utilizadas várias métricas diferentes:

- Distância euclidiana;
- Distância de Manhattan;
- Distância de Chebyshev;
- Outras.

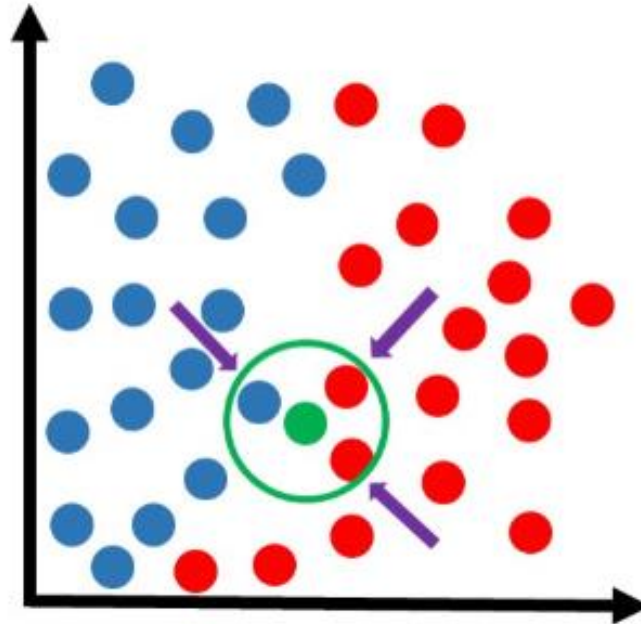
KNN

K = 3



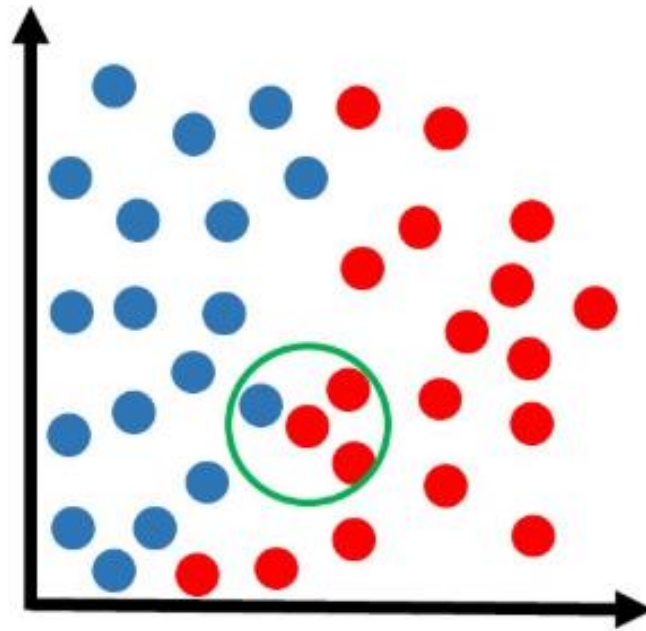
KNN

$K = 3$



KNN

K = 3





Obrigada!

Ana Laurentino



Deep learning

Deep learning

O termo *deep learning* vem do inglês e significa aprendizagem profunda e também é uma subárea de estudos da inteligência artificial. Sendo mais específico ainda, é uma subárea do aprendizado de máquina.

Nesta área é aprimorado o uso das redes neurais para aplicações de reconhecimento de fala, visão computacional e processamento de linguagem natural.

Essa dedicação específica se dá pelo fato dessas aplicações ainda não terem se desenvolvido tanto quanto outras e serem muito promissoras para o mercado.

Deep learning

Deep learning



Redes neurais

As primeiras redes neurais surgiram por volta de 1950 e se tornaram soluções importantes para aplicações em diversos setores do mercado empresarial.

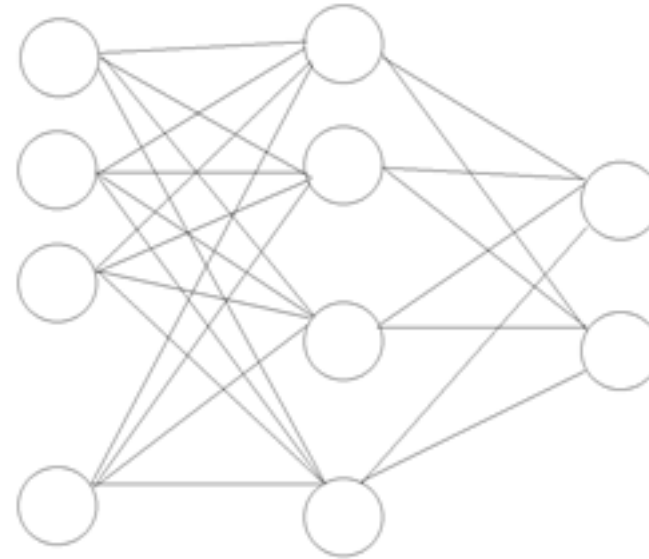
As redes neurais são técnicas que buscam simular o funcionamento dos neurônios humanos e as sinapses que realizam.

Com as redes neurais são construídos modelos e os neurônios são ajustados com pesos até que o modelo esteja satisfatório para a situação que representa.

Estrutura básica de uma rede neural

As redes neurais são compostas por, pelo menos:

- camada de entrada;
- camada escondida;
- camada de saída.



Como funciona o deep learning

No deep learning as redes neurais são modificadas para treinar redes neurais que possam aprender a processar linguagem natural.

Entende-se que dessa forma as redes neurais poderão desenvolver um aprendizado mais profundo sobre muitos problemas reais.

Como funciona o deep learning

Normalmente antes de aplicar algoritmos de aprendizado de máquina, os dados precisam ser preparados para estarem em uma estrutura que propicie o aprendizado.

O processo de tratamento dos dados é bastante complexo, exige profissionais especializados e demanda muito esforço computacional. Tudo isso sempre é realizado para cada base ou problema a ser trabalhada, ou seja, é uma solução muito custosa e específica.

Como funciona o deep learning

Nos algoritmos de aprendizado profundo, uma nova abordagem é proposta e diminui a quantidade de pré-processamento que é necessária para trabalhar com conjuntos de dados.

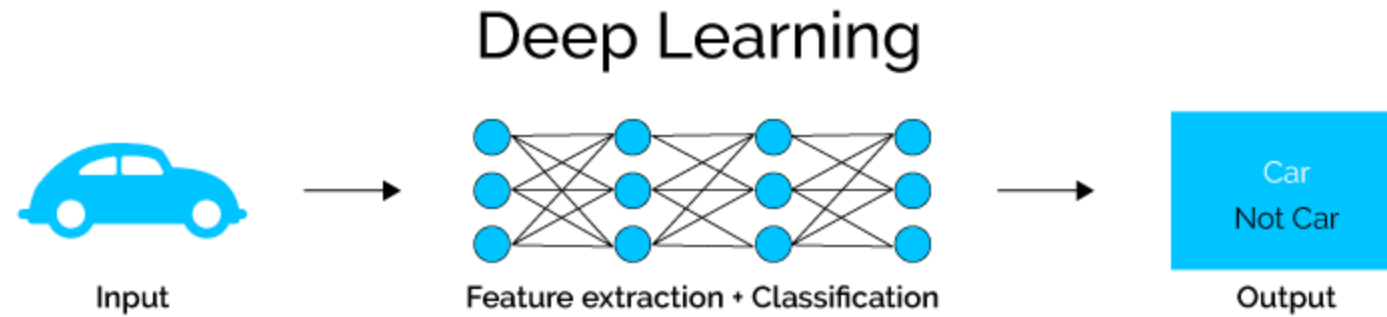
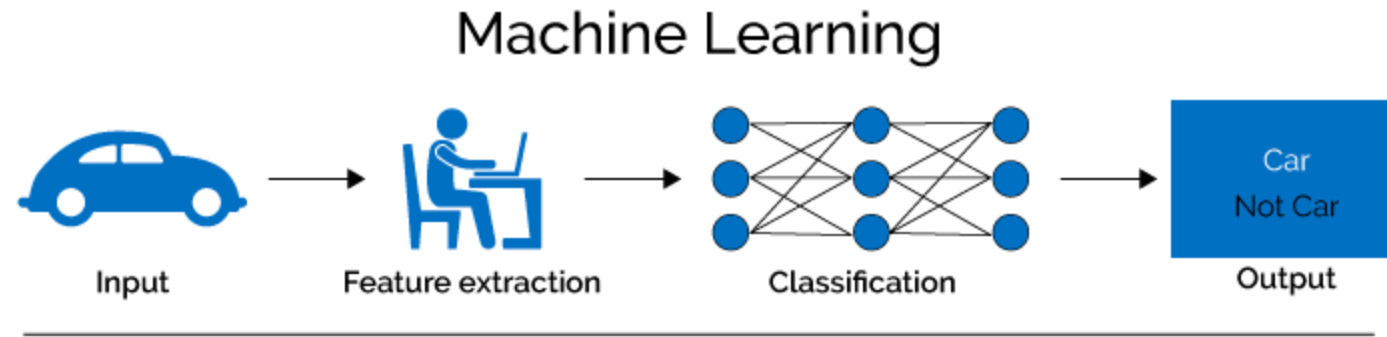
Dentro desses algoritmos são inseridos recursos para que as propriedades relevantes sejam automaticamente identificadas e tratadas em suas camadas hierárquicas.

Como funciona o deep learning

Por conta do comportamento citado, um mesmo algoritmo de aprendizado profundo pode ser utilizado em diferentes problemas e conjuntos de dados, apresentando um bom resultado. Isso porque as várias camadas não lineares atuam na obtenção de uma representação de qualidade.

A cada camada é atribuída uma funcionalidade, por exemplo, costuma-se tratar dados sensoriais como pixels na primeira camada. Ao sair da primeira camada, os dados alimentam automaticamente a camada posterior e assim por diante.

Deep learning x Machine learning



Onde está o deep learning

- As empresas Amazon e eBay usam para estudar o comportamento de seus clientes e consequentemente aumentar as suas vendas. Para isso os dados de compras e navegação são constantemente registrados;
- O Facebook utiliza o aprendizado profundo para realizar o reconhecimento facial em fotografias, sugerindo os usuários a serem marcados de acordo com quem aparece na foto;
- Entidades e empresas na área de segurança utilizam softwares de reconhecimento facial para identificar a presença de pessoas procuradas em eventos;
- Os chatbots utilizam o aprendizado profundo para realizar o processamento de linguagem natural e fazer o trabalho de atendimento ao consumidor.

Referências

SPERLING, Ed.; Deep Learning Spreads. 2018. Disponível em: <<https://semiengineering.com/deep-learning-spreads>>. Acesso em: 20 jan. 2020.



Obrigada!

Ana Laurentino



Obrigada!

Ana Laurentino

Algoritmo do Deep learning

O *deep learning* é baseado em redes neurais que vão conter muitas camadas para criar um modelo que trabalhe no tratamento e análise dos dados.

Embora todas as abordagens sejam baseadas em redes neurais, existem muitos modelos de redes neurais disponíveis com características específicas e, conseqüentemente, utilidades.

Redes neurais - Multilayer perceptron

O Multilayer Perceptron (MLP) é um algoritmo de rede neural bastante simples e que realiza uma classificação binária, ou seja, cada neurônio de saída emite 0 ou 1. Para classificar uma grande quantidade de classes é fundamental que se combine mais neurônios na camada de saída.

As redes MLP podem trabalhar com a classificação de problemas não lineares, ou seja, que possuem mais de uma classe, ao contrário do Perceptron puro.

Redes neurais - Multilayer perceptron

As redes neurais MLP com ao menos uma camada escondida pode aproximar funções contínuas.

Neste tipo de neural é obrigatória a presença de ao menos uma camada oculta ou escondida, além da camada de entrada e saída.

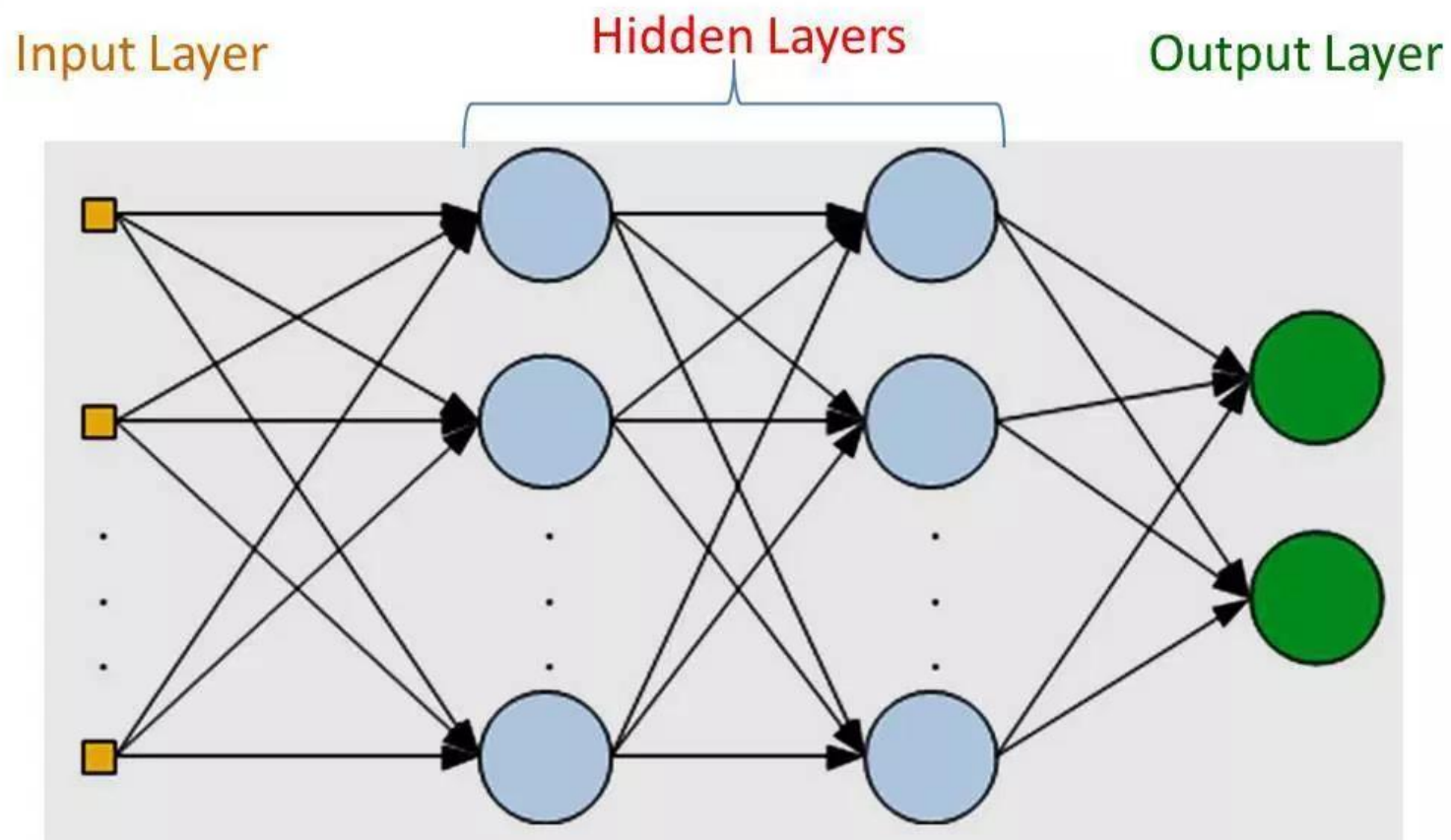
É uma abordagem bastante recomendada para quem está começando por conta de sua facilidade de compreensão e implementação.

Redes neurais - Multilayer perceptron

A abordagem MLP é utilizada para o aprendizado supervisionado, sendo treinada com dados de entrada e saída.

Para aprimoramento dos seus pesos é utilizado o algoritmo backpropagation que repassa o erro no sentido contrário da rede, distribuindo o erro por todos os neurônios.

Redes neurais - Multilayer perceptron



Redes neurais convolucionais

As redes neurais convolucionais são redes profundas que são especialmente interessantes para a classificação, agrupar por similaridade ou reconhecer objetos em imagens.

Com as redes neurais convolucionais é possível reconhecer caracteres e textos manuscritos e fazer a digitalização dos dados lidos em um arquivo.

Redes neurais convolucionais

A capacidade das redes convolucionais de fazer reconhecimento em imagem foi o que deu mais destaque ao aprendizado profundo.

As redes neurais trabalham as imagens com tensores que, basicamente, tem uma estrutura semelhante as matrizes.

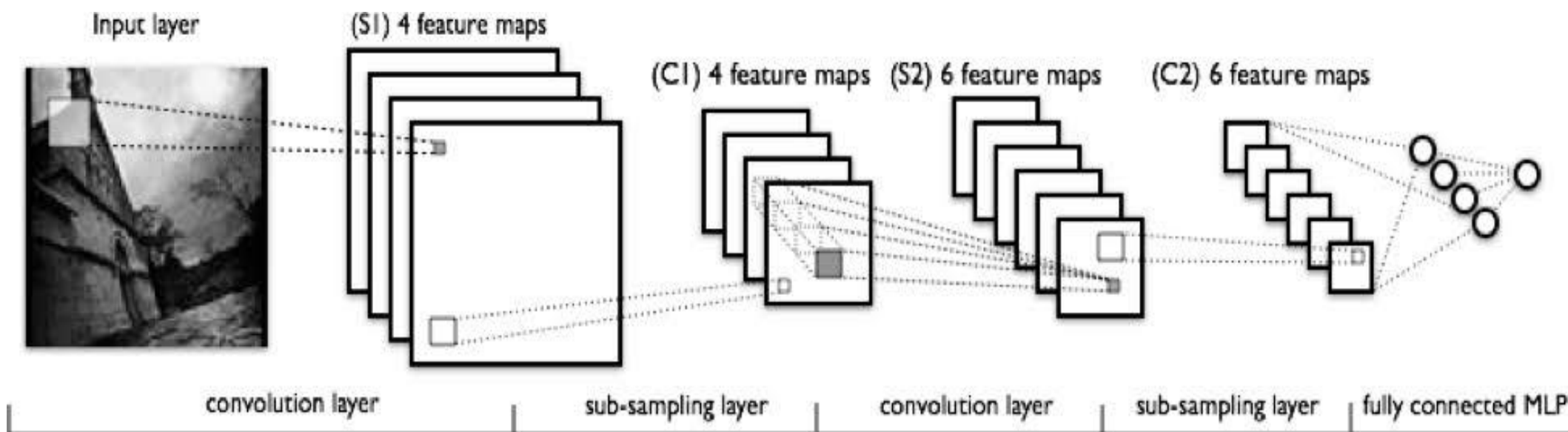
Redes neurais convolucionais - Imagens

Nas redes neurais convolucionais as imagens são recebidas em três estratos separados, um para cada intensidade das cores Red - Green - Blue (RGB).

O formato recebido é um retângulo com a quantidade de pixels que estavam presentes dentro da imagem.

A estrutura da imagem vai se modificando entre as camadas da rede neural, calculando uma probabilidade final para cada uma das possíveis saídas. A imagem será rotulada com a classe de maior probabilidade.

Redes neurais convolucionais



Redes neurais recorrentes

As redes neurais recorrentes se destacam no processamento de dados sequenciais como séries temporais e linguagem natural.

As redes recorrentes trabalham constantemente com um loop para feedback. Esse loop funciona pegando o resultado do passo $n-1$ para alimentar novamente a rede para que esse resultado possa auxiliar na construção do próximo passo.

Para ficar mais claro: se estiver recebendo um texto, ao ler a primeira letra do texto adquire-se um conhecimento importante para a leitura da próxima letra, e assim por diante.

Redes neurais recorrentes

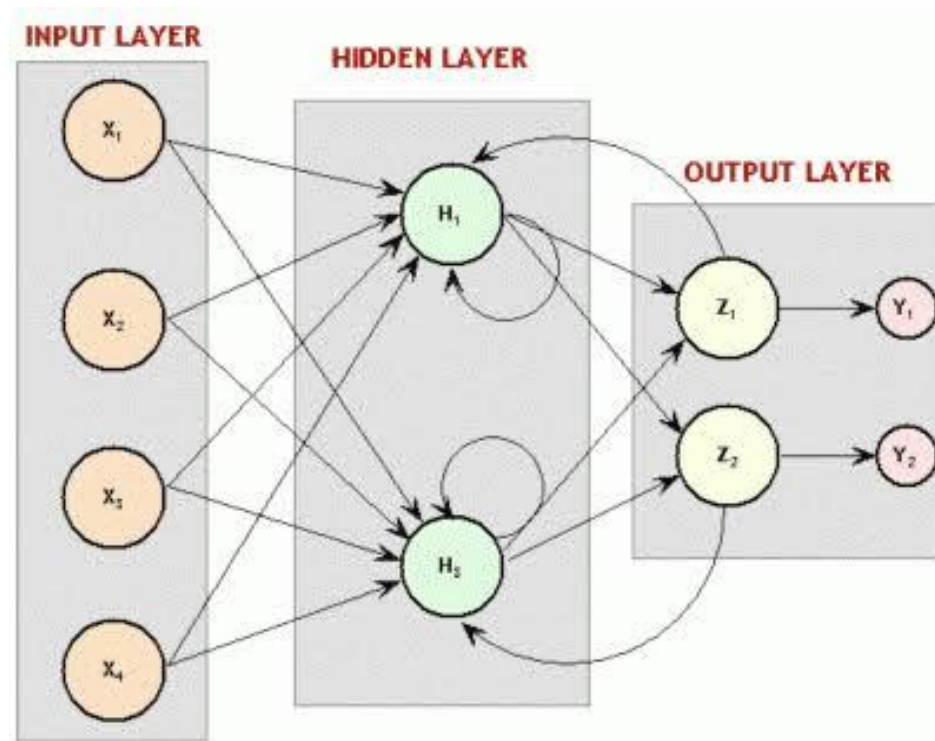
Os modelos das redes recorrentes tem como característica diferente a possibilidade de se transformar ao longo do tempo, aprendendo de acordo com o trabalho que vai realizando.

Outras redes para aprendizado profundo

Existem outras opções de rede para aprendizado profundo:

- Long Short-Term Memory;
- redes de Hopfield;
- Máquinas de Boltzmann;
- Deep Belief Network;
- Deep Auto-Encoders;
- Deep Neural Network Capsules.

Redes neurais recorrentes



Referências

RAVINDRA, S.; Como as Redes Neurais Convolucionais realizam o reconhecimento de imagem. 2017. Disponível em: <<https://www.infoq.com/br/articles/redes-neurais-convolucionais>>. Acesso em: 20 Jan. 2020.



Obrigada!

Ana Laurentino



Unyleya
EDUCACIONAL



Big Data

O conceito de Big Data

Com as pessoas trabalhando e se relacionando por meio da internet, a quantidade de dados disponível para exploração cresce progressivamente.

A detenção dos dados não é suficiente para garantir que eles serão bem aproveitados, isso porque a realização de análises manualmente se torna mais difícil à medida que o volume de dados aumenta.

O conceito de Big Data

O Big Data consiste em utilizar mecanismos computacionais para analisar um grande volume de dados.

Com a possibilidade de processar mais dados, a coleta e as fontes de dados também foram expandidas e as empresas passaram a coletar também dados de dentro da internet.

O conceito de Big Data

O Big Data é viabilizado por um conjunto de ferramentas que trabalham de forma coordenada para a coleta, análise e disponibilização de dados.

Relações do Big Data



O surgimento do Big Data

O armazenamento e utilização de dados é uma tarefa bastante antiga, antes mesmo de se concretizar o termo Big Data.

O termo Big Data foi consolidado no 2000, através de trabalhos do Doug Laney que estabeleceu o que seria essa técnica, a finalidade e como colocá-la em prática.

Os 5 V's

O Big Data propõe gerar informações de valor para dentro das empresas considerando diferentes variáveis, que são os chamados 5 V's:

- Volume;
- Velocidade;
- Variedade;
- Variabilidade;
- Vínculo.

Volume

O volume é talvez um dos pontos centrais do Big Data, porque o seu surgimento se deu pela necessidade de tratar grandes quantidades de dados.

Então presume-se que uma aplicação de Big Data será constituída a partir de um grande agrupamento de dados.

Outras tecnologias podem ser associadas para facilitar o armazenamento e manipulação de grandes quantidades de dados, como os bancos de dados NoSQL e o armazenamento de dados nas nuvens.

Velocidade

Para lidar com grandes quantidades de dados em tempo hábil, as velocidades de transferência e processamento de dados precisa ser grande.

Todo o processamento e transações do Big Data precisam ser bem otimizados para que a tarefa não se torne um problema, ao invés de solução.

Variedade

Em aplicações de processamento de dados mais tradicionais, o mais comum é que se use somente com uma fonte de dados para serem trabalhados.

No caso do Big Data, a variedade é estimulada, tanto do ponto de vista da fonte quanto da estrutura. Mas isso exige que os dados sejam trabalhados e adequados para o processamento.

O interesse na variedade é permitir que novos pontos de vista sejam construídos e se formem melhores informações ao final das análises.

Variabilidade

A produção de dados que são utilizados para o Big Data, por exemplo dentro da internet, varia em estrutura e na velocidade com que são produzidos.

Dentro da internet, os dados são gerados sem um controle das empresas, eles são produzidos de acordo com os interesses dos usuários.

Vínculo

Ao absorver grandes quantidades de dados, é fundamental que se crie uma estrutura para organizá-los e permitir que eles sejam acessados sempre que for necessário.

Imagine possuir um grande volume de dados sem qualquer organização: como seria possível encontrar aquilo que precisa de forma rápida?

Dados estruturados

Os dados estruturados são aqueles que estão organizados em uma formato conhecido e fácil de ser explorado pelas aplicações do Big Data.

Esses dados podem ser dados log, resultados de formulário, entre outros.

Dados não estruturados

Os dados não estruturados são aqueles que não possuem um padrão de organização, contendo recursos dispostos em diferentes formatos.

Isso acontece especialmente porque os dados são de diferentes fontes, e mesmo em uma das fontes eles podem variar.

Referências

PODEROSO, C.; Visão geral da arquitetura de Big Data. 2014. Disponível em: <<https://canaltech.com.br/big-data/Visao-geral-da-Arquitetura-de-Big-Data>>. Acesso em: 20 Jan. 2020.



Obrigada!

Ana Laurentino