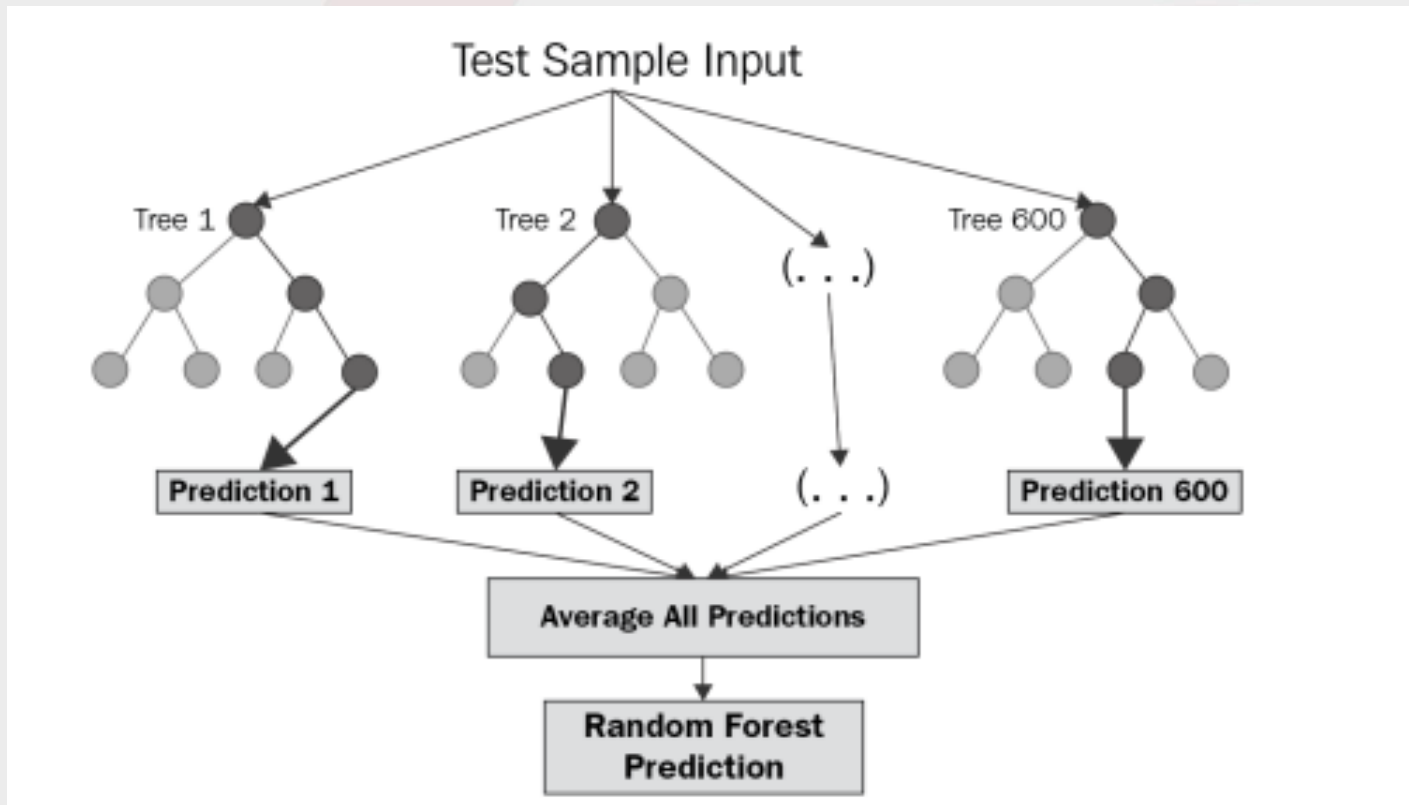


Random Forest

- Random Forest é um algoritmo de classificação baseado em árvore de decisão.



Random Forest

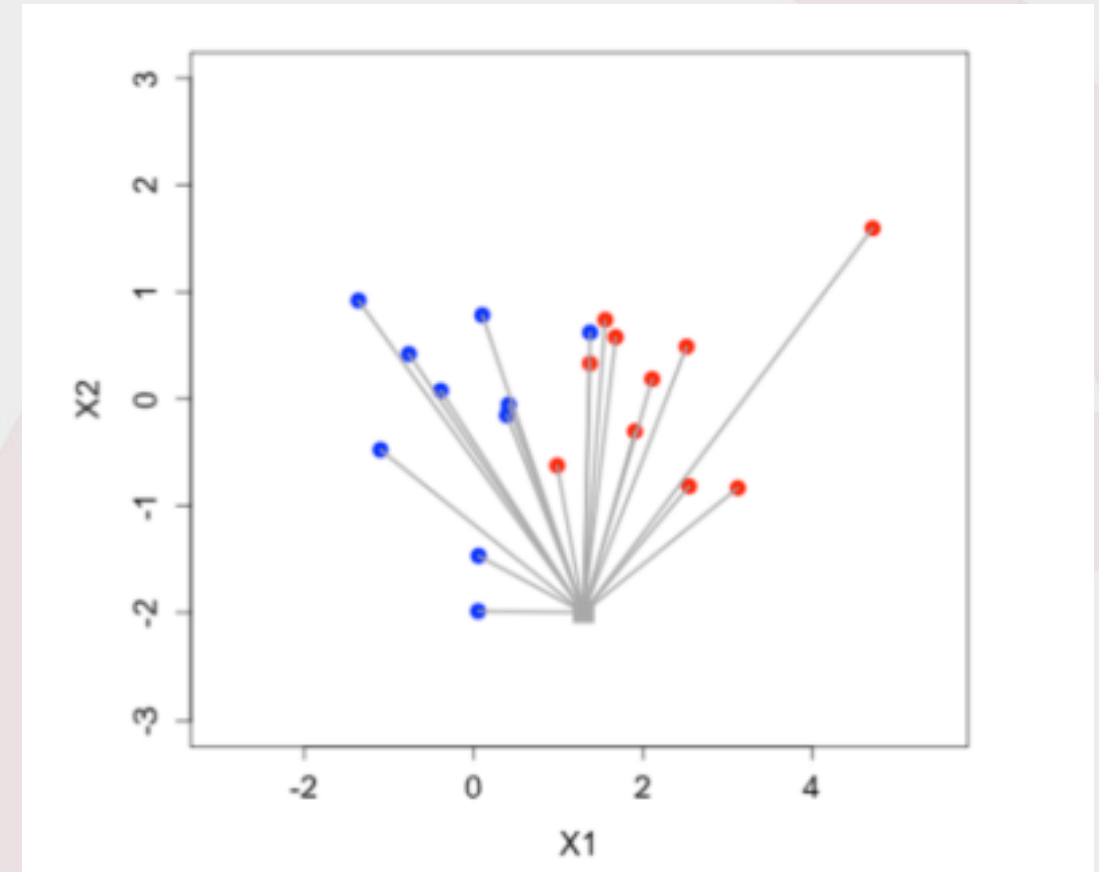
- Prova: Quadrix - 2022 - CRA-PR - Analista Sistemas I
- Com relação aos conceitos de machine learning e deep learning, julgue o item.
- Entre as técnicas de machine learning, a random forest é capaz de solucionar problemas de classificação e de regressão, por meio da construção e dos treinamentos de árvores de decisão.

Random Forest

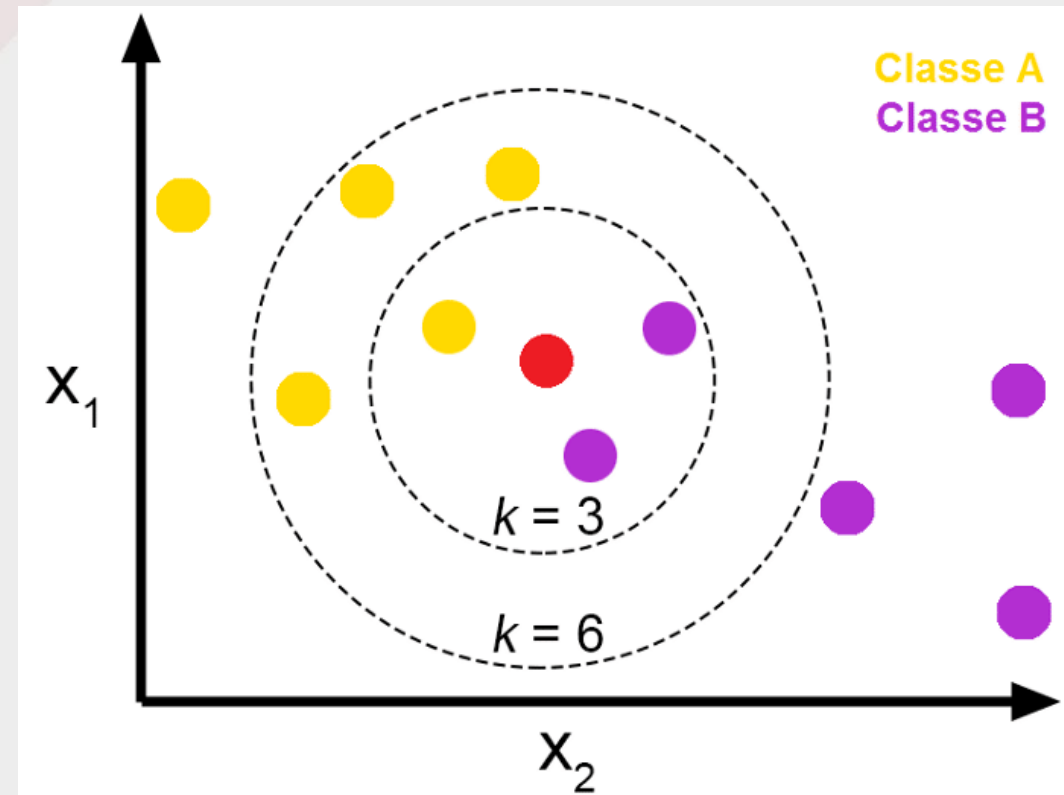
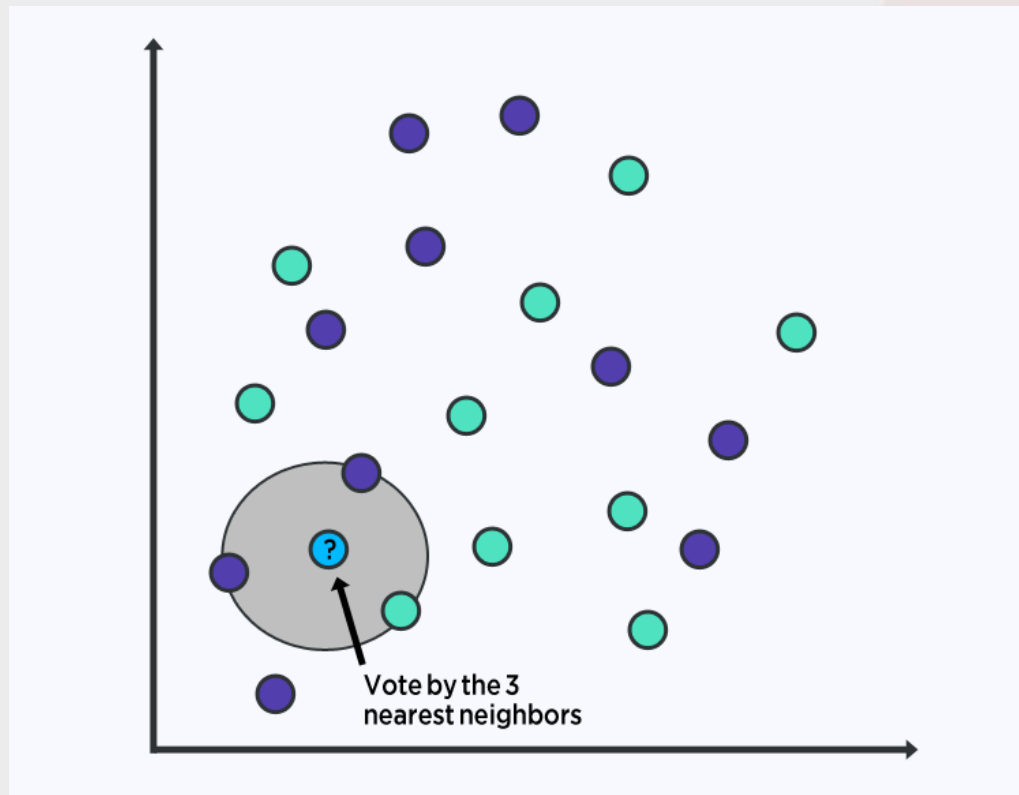
- Prova: Quadrix - 2022 - CRA-PR - Analista Sistemas I
- Com relação aos conceitos de machine learning e deep learning, julgue o item.
- Entre as técnicas de machine learning, a random forest é capaz de solucionar problemas de classificação e de regressão, por meio da construção e dos treinamentos de árvores de decisão.

k-Nearest Neighbors

- KNN é um algoritmo que classifica novos dados com base em uma medida de similaridade entre seus "vizinhos" mais próximos, ou seja, aqueles que têm características semelhantes às suas.
- Nesse método, utiliza-se a distância (usualmente a Euclidiana) entre uma nova observação e as demais observações de um training set para classificá-la de acordo com a observação mais próxima.
- O "k" determina o número de vizinhos que serão utilizados para a classificação.



k-Nearest Neighbors



KNN – Questões de Concurso

- Prova: CESGRANRIO - 2021 - Banco do Brasil - Agente de Tecnologia
- Um pesquisador conseguiu uma base de dados que mostrava terrenos classificados de acordo com:
 - características físicas;
 - tipo de negócio a ser nele implantado;
 - risco esperado, que compreendia os rótulos alto, médio, baixo ou nenhum.
- Decidiu, então, usar um algoritmo de aprendizado de máquina que, a partir das características físicas do terreno e do tipo de negócio a ser nele implantado, aprenderia a determinar o risco esperado, enquadrando o terreno em questão em um daqueles rótulos.
- Nesse cenário, que algoritmo de aprendizado de máquina é indicado para resolver esse problema?
- A PCA
- B K-NN
- C DBSCAN
- D K-Medoids
- E Redes de Kohonen

KNN – Questões de Concurso

- Prova: CESGRANRIO - 2021 - Banco do Brasil - Agente de Tecnologia
- Um pesquisador conseguiu uma base de dados que mostrava terrenos classificados de acordo com:
 - características físicas;
 - tipo de negócio a ser nele implantado;
 - risco esperado, que compreendia os rótulos alto, médio, baixo ou nenhum.
- Decidiu, então, usar um algoritmo de aprendizado de máquina que, a partir das características físicas do terreno e do tipo de negócio a ser nele implantado, aprenderia a determinar o risco esperado, enquadrando o terreno em questão em um daqueles rótulos.
- Nesse cenário, que algoritmo de aprendizado de máquina é indicado para resolver esse problema?
- A PCA
- B K-NN
- C DBSCAN
- D K-Medoids
- E Redes de Kohonen

KNN – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2021 - SEFAZ-AL - Auditor Fiscal de Finanças e Controle de Arrecadação da Fazenda Estadual
- O método k-NN (k-nearest neighbors) é um dos classificadores não paramétricos baseados em distância.

KNN – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2021 - SEFAZ-AL - Auditor Fiscal de Finanças e Controle de Arrecadação da Fazenda Estadual
- O método k-NN (k-nearest neighbors) é um dos classificadores não paramétricos baseados em distância.

KNN – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Com respeito a métodos para imputação de dados, julgue o seguinte item.
- O método de imputação K-NN (k-nearest neighbours) leva em consideração os padrões de similaridade presentes no conjunto de dados para prever os valores faltantes. No entanto, a escolha da função de distância para a aplicação desse método, como, por exemplo, HEOM (heterogeneous euclidean-overlap metric) ou HVDM (heterogeneous value difference metric), pode influenciar significativamente nos resultados da imputação.

KNN – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Com respeito a métodos para imputação de dados, julgue o seguinte item.
- O método de imputação K-NN (k-nearest neighbours) leva em consideração os padrões de similaridade presentes no conjunto de dados para predizer os valores faltantes. No entanto, a escolha da função de distância para a aplicação desse método, como, por exemplo, HEOM (heterogeneous euclidean-overlap metric) ou HVDM (heterogeneous value difference metric), pode influenciar significativamente nos resultados da imputação.

Naive Bayes

- Família de algoritmos de aprendizado supervisionado baseados em probabilidade.
- Teorema de Bayes:



$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Diagram illustrating the components of Bayes' Theorem:

- $P(A|B)$: THE PROBABILITY OF "A" BEING TRUE GIVEN THAT "B" IS TRUE
- $P(B|A)$: THE PROBABILITY OF "B" BEING TRUE GIVEN THAT "A" IS TRUE
- $P(A)$: THE PROBABILITY OF "A" BEING TRUE
- $P(B)$: THE PROBABILITY OF "B" BEING TRUE

Naive Bayes

Exemplo

- Um médico sabe que a meningite causa torcicolo em 50% dos casos. Porém, o médico sabe que a meningite atinge 1/50.000 e também que a probabilidade de se ter torcicolo é de 1/20.
- Usando Bayes para saber a probabilidade de uma pessoa ter meningite dado que ela está com torcicolo

Temos então

$$\bullet \quad P(T|M) = 0,5 \quad P(M) = 1 / 50.000 \quad P(T) = 1 / 20$$

$$P(M|T) = \frac{P(M) \times P(T|M)}{P(T)} = \frac{\frac{1}{50000} \times 0.5}{1/20}$$

$$P(M|T) = 0,0002 = \frac{2}{10000} = 0,02\%$$

Naive Bayes

- Suposições ingênuas:
 - Todas as características do conjunto de dados não dependem um do outro.
 - Cada evento contribui igualmente para classificar o resultado.



Naive Bayes

- Mais características:
 - É rápido -> previsão em tempo real.
 - Famoso como classificador de spam, de sentimentos de recomendações.
- Funcionamento:
 - Prevê uma tabela de probabilidades condicionais e depois calcula a saída com base nessa tabela.



Naive Bayes

- Prova: CESPE / CEBRASPE - 2021 - SEFAZ-CE - Auditor Fiscal de Tecnologia da Informação da Receita Estadual
- A classificação Naive Bayes parte da suposição de que as variáveis envolvidas em machine learning são independentes entre si.

Naive Bayes

- Prova: CESPE / CEBRASPE - 2021 - SEFAZ-CE - Auditor Fiscal de Tecnologia da Informação da Receita Estadual
- A classificação Naive Bayes parte da suposição de que as variáveis envolvidas em machine learning são independentes entre si.

Naive Bayes

- Prova: CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados
- Devido ao fato de pressupor independência entre atributos, o algoritmo Naive Bayes é capaz de realizar, com precisão, o treinamento de um modelo com uma quantidade reduzida de amostras.

Naive Bayes

- Prova: CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados
- Devido ao fato de pressupor independência entre atributos, o algoritmo Naive Bayes é capaz de realizar, com precisão, o treinamento de um modelo com uma quantidade reduzida de amostras.

Naive Bayes

- Prova: INSTITUTO AOCP - 2020 - MJSP - Cientista de Dados - Big Data
- Um cientista de dados deve utilizar uma técnica de mineração de dados que não usa regras ou árvore de decisão ou qualquer outra representação explícita do classificador, mas que usa o ramo da matemática conhecido como teoria das probabilidades para encontrar a mais provável das possíveis classificações. Sabendo disso, assinale a alternativa que apresenta corretamente o nome da técnica a ser utilizada pelo cientista de dados.
- A Regras de decisão.
- B Árvore binária.
- C Entropia.
- D Classificação Naive Bayes.
- E Agrupamento.

Naive Bayes

- Prova: INSTITUTO AOCP - 2020 - MJSP - Cientista de Dados - Big Data
- Um cientista de dados deve utilizar uma técnica de mineração de dados que não usa regras ou árvore de decisão ou qualquer outra representação explícita do classificador, mas que usa o ramo da matemática conhecido como teoria das probabilidades para encontrar a mais provável das possíveis classificações. Sabendo disso, assinale a alternativa que apresenta corretamente o nome da técnica a ser utilizada pelo cientista de dados.
- A Regras de decisão.
- B Árvore binária.
- C Entropia.
- D Classificação Naive Bayes.
- E Agrupamento.