

CIÊNCIA DA COMPUTAÇÃO – 7º Período

RELATÓRIO – 1 – INTELIGÊNCIA ARTIFICIAL Prof. Fischer Stefan

- 1) Explicar o algoritmo KNN. Considere em sua explicação sobre o fato de ser um método supervisionado ou não; de se conhecer rótulo ou não e o que exatamente ele executa.

O método KNN consiste em classificar uma certa amostra de dados com base nos valores já conhecidos, sendo definida em relação aos vizinhos mais próximos.

Passos principais do KNN:

- 1- Recebe um conjunto de dados já classificados e com rótulo definido.
- 2- Definição do número de vizinhos mais próximos que serão utilizados para classificar o novo ponto de dados desconhecido.
- 3- Calcula a distância do novo dado a todos os outros dados do conjunto.
- 4- Identifica os vizinhos mais próximos.
- 5- Define a classificação do novo dado, caso for uma tarefa de classificação define pela classe mais comum entre os vizinhos mais próximos, e caso for uma tarefa de regressão, o valor será definido pela média dos vizinhos mais próximos.

Este método é supervisionado, tendo como definição um sistema que é programado a partir de um conjunto de dados já estabelecidos, e utilizando estes dados para tomar decisões quando recebe novos dados.

Os rótulos são extremamente importantes, já que são cruciais para classificar novos pontos de dados, com base nas informações aprendidas durante o treinamento, sem eles o algoritmo não teria capacidade de realizar a classificação.

- 2) Considerando o exercício discutido em sala, qual o papel da coluna TARGET CLASS?

target class é o resultado desejado que queremos que o modelo preveja ou classifique corretamente durante a fase de treinamento e, posteriormente, ao lidar com novos dados durante a fase de teste ou inferência.

- 3) Por que é necessário padronizar os dados?

A padronização dos dados é de suma importância, visto que se os valores não estiverem em escalas similares não é possível fazer a comparação dos mesmos, dando maior exatidão e facilitando as comparações feitas.

4) Explicar o método `scaler.fit()`

O método `scaler.fit()` é importante para preparar o escalador em relação aos dados de treinamento, tendo a função de padronizar e normalizar os dados antes do treinamento do modelo de aprendizado de máquina.

5) Explicar a execução do `scaled_features` e `def_feat`:

`Scaled_Features`: Esta função basicamente faz com que todas as classes e características fiquem na mesma escala, garantindo que não dê preferência para uma certa característica por ela ter valores maiores, normalizando e padronizando os dados.

`Def_feat`: Esta função basicamente filtra as características antes de serem utilizadas, tanto as que não são úteis quanto as que precisam ser pré-processadas antes de serem usadas, restando apenas aquelas que são relevantes.

Ambas são etapas importantes no pré-processamento de dados antes de treinar nosso modelo de aprendizado de máquina.

6) Explicar o resultado `Classification Report`

Quando você tem um modelo que tenta prever diferentes categorias para seus dados, o "Classification Report" ajuda a entender o quão bem esse modelo está indo em cada categoria.

1. Precision:

- Precisão é sobre o quão certo o modelo é quando ele diz que uma amostra é de uma determinada categoria. Se ele diz que algo é um "gato", a precisão nos diz quantas vezes ele está certo.

- Por exemplo, se ele diz que há 10 gatos, mas apenas 8 são realmente gatos, então a precisão é $8/10 = 0,8$ (ou 80%).

2. Recall:

- Revocação é sobre o quão bem o modelo consegue encontrar todas as amostras de uma categoria.

- Por exemplo, se há 10 gatos no total, mas o modelo só encontra 8 deles, então a revocação é $8/10 = 0,8$ (ou 80%).

3. F1-Score:

- O F1-Score é uma maneira de combinar precisão e revocação em uma única métrica. Ele nos dá uma ideia geral de quão bem o modelo está fazendo.
- É útil porque leva em conta tanto os falsos positivos quanto os falsos negativos (quando o modelo erra em dizer que algo é de uma categoria quando na verdade não é, e vice-versa).

4. Support:

- Suporte é apenas o número real de amostras que pertencem a cada categoria. É uma espécie de contexto para as outras métricas.

Basicamente, o "Classification Report" nos dá uma visão de como o modelo está se saindo em cada categoria, ajudando a entender se ele está tendo problemas com alguma categoria específica e onde pode precisar melhorar. É como um relatório que nos dá uma ideia de onde focar nossos esforços para fazer o modelo funcionar melhor.

7) Diferenciar as técnicas de Regressão Linear e Logística

Regressão Linear:

- A Regressão Linear é usada quando queremos prever um valor numérico contínuo com base em uma ou mais variáveis independentes.
- Por exemplo, se quisermos prever o preço de uma casa com base em sua área, número de quartos, etc., podemos usar a Regressão Linear.
- A ideia é encontrar uma linha reta (ou um hiperplano, em casos com múltiplas variáveis) que melhor se ajuste aos dados, minimizando a soma dos quadrados das diferenças entre os valores previstos pela linha e os valores reais.
- O objetivo é encontrar a relação linear entre as variáveis independentes e a variável dependente (o valor que queremos prever).

Regressão Logística:

- A Regressão Logística é usada quando queremos prever a probabilidade de uma determinada categoria ou evento ocorrer, com base em uma ou mais variáveis independentes.

- Por exemplo, se queremos prever se um paciente tem uma determinada doença com base em seus sintomas, podemos usar a Regressão Logística.

- Em vez de prever um valor contínuo, como na Regressão Linear, a Regressão Logística prevê a probabilidade de uma categoria específica. Essa probabilidade é então convertida em uma decisão de classe (por exemplo, "sim" ou "não", "doente" ou "saudável") usando um limite de decisão.

- A função logística (ou sigmoide) é usada para modelar a relação entre as variáveis independentes e a variável dependente. Essa função garante que as saídas do modelo estejam entre 0 e 1, representando probabilidades.

- O objetivo é encontrar os coeficientes que melhor se ajustem aos dados, maximizando a verossimilhança dos dados observados sob o modelo.

Resumindo, enquanto a Regressão Linear é usada para prever valores contínuos, como preços ou quantidades, a Regressão Logística é usada para prever probabilidades de classes, como sim/não ou doente/saudável. Cada uma tem suas aplicações específicas dependendo do tipo de problema que estamos tentando resolver.