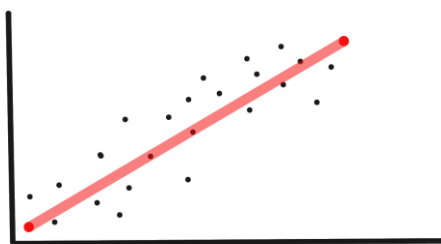


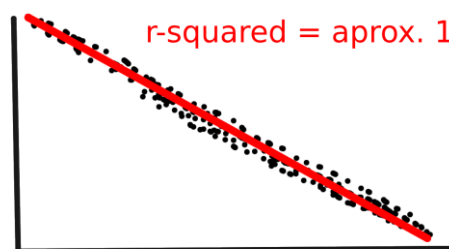
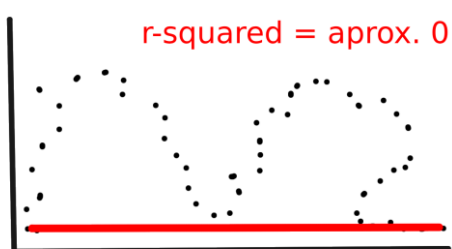
Predictive Models

Linear Regression

Regressão linear é o primeiro dos modelos utilizados para prever valores, sua tarefa é criar uma linha reta dentro do gráfico para mostrar o trajeto destes dados, com essa reta conseguimos identificar caso o gráfico esteja crescendo, decrescendo ou em constância.

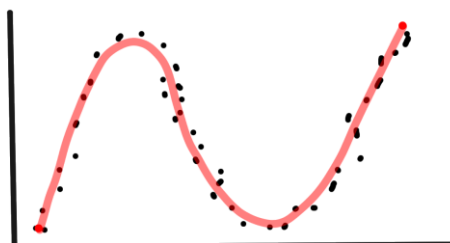


Para descobrirmos o quão eficaz a regressão linear será no gráfico, podemos calcular o seu r-squared, sendo um valor que varia de 0 até 1 (sendo 1 um modelo perfeito para ser utilizado).



Polynomial Regression

Considerando que nem todos os gráficos possuem valores lineares, podemos utilizar da regressão polinomial, onde podemos aumentar a ordem conforme a complexidade aumenta.



Um problema enfrentado com este método é a sobrecarga de ordem, na qual o modelo pode estar utilizando um polinômio de grau maior do que o necessário, causando problemas tanto de otimização como de acurácia.

O uso do r-squared tem sua utilidade, mas não é um bom método para encontrar essa utilização excessiva, por isso é utilizado o train/test.

Multiple Regression

Em casos de possuir mais que uma variável que influencie o que você está interessado. Por exemplo, imagine que você está comprando um monitor, você não consideraria apenas o seu tamanho, mas sim as especificações presentes nele, resolução, tipo de painel, garantia etc. Sendo cada atributo desse responsável por modificar o seu valor final.

Com esse método é possível separar uma importância para cada variável, eliminando assim as com menor importância geral.

Podemos facilitar esse trabalho simplesmente por pesquisar a diferença de partes em específico, melhorando a eficiência para projetos de escala menor. Utilizando o exemplo do monitor podemos comparar os preços médios de monitores 1080p com monitores 1440p, retornando que aumentar a resolução possui uma influência no seu valor.

Multi-Level Models

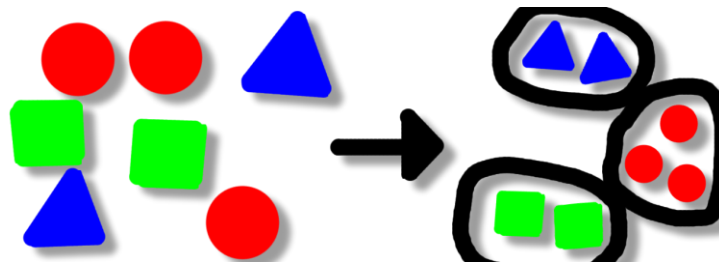
É o conceito que mostra como diferentes efeitos ocorrem em diferentes níveis. Por exemplo, a sua educação pode ter uma influência sobre o modo que você foi criado, o ambiente em que vive, a cidade em que vive, a sua riqueza etc. O modelo de multi-level tenta trazer todos esses aspectos para o cálculo final.

Machine Learning with Python

Unsupervised Learning

Aprendizado não supervisionado acontece quando o modelo que você está treinando não recebe nenhuma "resposta" sobre aquilo que ele está aprendendo, ele cria um sentido baseado no grupo de dados que foi fornecido.

Por exemplo, tentar organizar grupos de objetos baseados em seus formatos, o modelo não saberá o que faz eles serem parecidos ou o que utilizar como parâmetro.



Supervised Learning

No aprendizado supervisionado, o modelo utiliza das "respostas" que possui para "aprender" como separar os dados com suas respectivas funções.

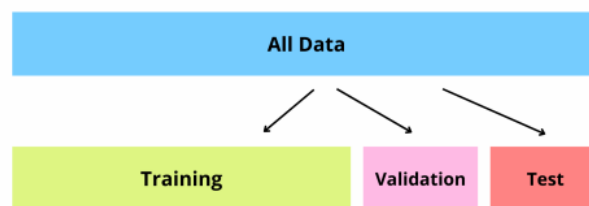
Esse modelo é então utilizado para prever respostas para outros cenários.

Por exemplo, é possível prever o preço de um carro que ainda não existe baseando-se nos valores de modelos passados, partindo do ano, modelo, preço de lançamento etc.

$$X + 20 = ??? \quad \left\{ \begin{array}{l} X + 1 = 6 \\ \text{Logo,} \\ X + 20 = 25 \end{array} \right.$$

Train/Test

É o princípio de treinar e testar o seu modelo, considerando que possua muitos dados, é separado uma parte para o treinamento e o restante para os testes, fazendo assim com que seu modelo esteja funcional para diversas possibilidades futuras.



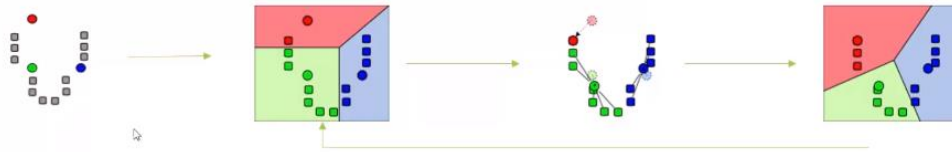
K-Means Clustering

Tenta dividir os dados em grupos K que estão mais próximos do centro K. Utiliza do aprendizado não-supervisionado focando apenas na posição de cada dado.

O modo como funciona é simples.

- Escolhe um centro K de forma aleatória.

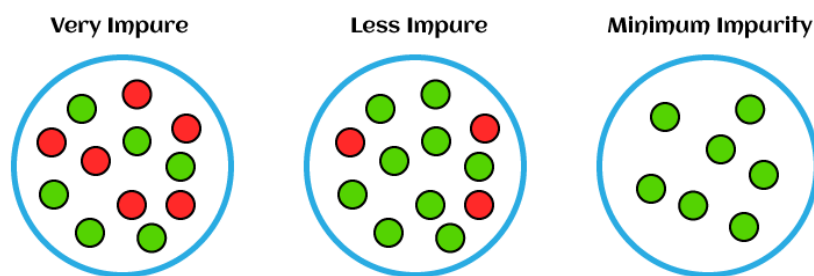
- Alinha cada dado para o K mais perto.
- Refaz o trabalho do centro baseado na posição de cada dado.
- Repete até os pontos se estabilizarem.



Entropy

A medida dos dados que possui diferença, o quanto igual ou diferente os dados são.

Por exemplo, se classificarmos animais pelos atributos e espécies, podemos considerar que num grupo de dados onde todos os animais são cachorros, sua entropia é equivalente a 0, pois todos são iguais em espécie.



Decision Trees

Outra forma de aprendizado supervisionado, é construído uma "árvore" na qual ajuda a decidir quais ações tomar, sendo o topo da árvore o resultado de todos os caminhos tomados.

Por exemplo, considere que você possua um histórico de dados das pessoas contratadas numa empresa, é possível construir uma árvore se baseando nesses dados para treinar o modelo e chegar no sistema que prevê qual será o candidato contratado.



Ensemble Learning

Aprendizado por agrupamento se baseia na ideia de combinar vários modelos para "votar" em qual será o melhor resultado.

Por exemplo, caso o resultado do algoritmo que você está treinando não esteja com uma precisão muito alta, talvez compense utilizar de diversos algoritmos para comparar e decidir qual o melhor modo de lidar com o problema.

XGBoost

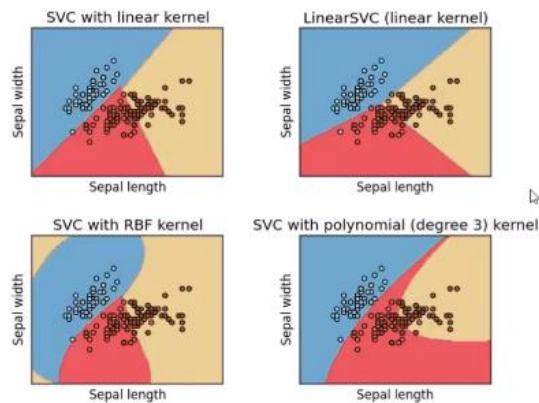
Um dos algoritmos mais utilizados por cientistas de dados, apresentando resultados superiores em problemas de previsão.

Ele é capaz de prever o overfitting, lidar com dados faltantes de forma automática, medir a performance a cada iteração, ser melhorado após sua implementação entre outras coisas.

Support Vector Machines (SVM)

Funciona bem para classificar dados de dimensões maiores.

É um conjunto de métodos de aprendizado supervisionado que analisam os dados e reconhecem padrões, utilizado para análise de regressão.



Atividade

Ao fim deste trabalho foi criado um código para exemplificar uma parte do conhecimento adquirido, a parte escolhida foi sobre a árvore de decisões.

Com bastante semelhança ao apresentado no vídeo, este código criado apresenta um modelo que decide caso o carro usado vale ou não a pena ser comprado, ele leva em consideração alguns fatores como preço, ano, portas e outros.

O código após ser treinado se mostra eficaz em "adivinhar" a resposta final, possuindo semelhanças e diferenças ao demonstrado do vídeo, é uma forma de mostrar aquilo que foi aprendido em prática com possíveis dados reais.

Conclusão

O documento apresentou diversos modelos preditivos usados em análise de dados e aprendizado de máquina, incluindo Regressão Linear, Regressão Polinomial, Regressão Múltipla e Modelos de Múltiplos Níveis. Foram também discutidos conceitos de Aprendizado Supervisionado e Não Supervisionado, com destaque para técnicas como K-Means Clustering, Árvores de Decisão, Support Vector Machines, Ensemble Learning e XGBoost.

A prática de Train/Test foi enfatizada para validar a eficácia dos modelos. Em resumo, o uso adequado dessas técnicas permite criar sistemas preditivos eficientes, que oferecem insights valiosos e suportam decisões informadas em vários contextos.