DCE44

__/__/__

02 - Aprendizado de Maq. Não - Supervisionado

Os modelos de regressão linear ou redes neurais utilizam um conjunto de parâmetros para prever a saída de uma determinada hipotese. Tais parâmetros são ajustados na etapa de treinamento de acordo com o que é esperado na saída de modo a obter-se um menor erro possível entre o valor predito e o valor real. Neste caso, há uma supervisão na etapa de treinamento visando os ajustes dos parâmetros. Portanto, temos um modelo paramétrico e de aprendizado supervisionado.

Quando se tem milhões ou bilhões de dados no conjunto de treinamento (Dataset), uma abordagem que pode trazer excelentes resultados seria construir um modelo onde os dados "falem" por si e não por intermedio de vetores parametrizados que muitas vezes são relativamente pequenos para o problema. Além disso, vale ressaltar que para modelos parametricos o número de parâmetros não muda com o tamanho do dataset de treinamento. Se for considerado uma hipotese que faz uso de todos os dados de treinamento para prever uma saída, tem-se um modelo não paramétrico e de aprendizagem não-supervionada. Nesse caso, não há ajuste de parâmetros e, claramente, não há nenhuma supervisão, o treinamento consiste o dataset sem nenhum ajuste orientado a um valor esperado.

Um exemplo bem simples, e com viés didático, seria inserir todos os dados de treinamento em uma tabela. Quando for necessário fazer a previsão da hipotese ($h(x)$)

__/__/____

o valor x pode ser consultado na tabela e retornar o y correspondente. Entretanto, tal proposta demonstra claramente uma dificuldade de generalização, pois caso x não exista na tabela, um valor padrão é retornado. Note que no modelo proposto não houve ajuste de parametros de forma supervisionada.

O aprendizado de maquina não-supervisionado possui caracteristicas extremamentes relevantes no contexto de lidar com muitos dados de entrada, pois a etapa de treinamento desses modelos requer menos tempo e menor esforço computacional. Além disso, tais modelos possuem maior facilidade na definição dos ~~hiperparametros~~ ~~etros~~ hiperparametros pelo profetista devido a ser um modelo menos complexo em geral. Por outro lado, tais modelos possuem dificuldades de lidar com generalizações e problemas complexos.

Um classico exemplo de modelos de aprendizagem não-supervisionada é o modelo K-means. Para caso de classificação, onde o objetivo é atrelar a(s) entrada(s), ou hipotese($h(x)$), a um certo grupo (saída), o modelo K-means faz a análise dos K pontos mais próximos de x, usando distância Euclidiana. Neste caso, é como se cada ponto selecionado votasse que x pertence ao seu ~~ponto~~ grupo (grupo do ponto selecionado). Por fim, x será classificado ao grupo que receber mais votos.

Para casos de regressão, onde a saída pode ser valores do dominio real, é possivel definir regressões lineares a partir dos centroides definidos ~~pelo mo~~ dos grupos definidos pelo modelo

SD

__/__/____

~~[redacted]~~ K-means.