



UFOP

Universidade Federal
de Ouro Preto

Estudo Dirigido 03

**Aluno em Graduação da Universidade
Federal de Ouro Preto do curso Ciência da**

Computação:

Halliday Gauss Costa dos Santos.

Matrícula: 18.1.4093.

Área: Inteligência Artificial.

Questão 1)

Na regressão logística, utiliza-se uma probabilidade para saber o rótulo de uma instância que será classificada, e esse rótulo pode ser 0 ou 1. Se o valor da predição for negativo, têm-se uma probabilidade menor que 0.5, porém, se o valor de predição for maior que 0, têm-se uma probabilidade maior que 0.5, e através dessas probabilidades, é possível calcular o valor do rótulo. Se a probabilidade for inferior a 0.5, então o rótulo y assumirá o valor 0, no entanto, se essa probabilidade for superior a 0.5, o rótulo y será igual a 1.

Com o rótulo e a probabilidade calculada, é possível encontrar a função custo, que é representada pela seguinte fórmula:

$$c(\theta) = \begin{cases} -\log(\hat{p}) & \text{se } y = 1 \\ -\log(1 - \hat{p}) & \text{se } y = 0 \end{cases}$$

Em que “ p ” é o valor da probabilidade da classificação e y é o rótulo classificado. Com essa fórmula, têm-se duas possibilidades para o cálculo do custo, que será escolhida de acordo com o valor do rótulo. Quando o rótulo é igual a 1, é utilizado a expressão $-\log(\text{probabilidade})$ como cálculo do custo, já quando o rótulo assume o valor 0, utiliza-se como expressão do custo $-\log(1-\text{probabilidade})$.

Questão 2)

A Regressão Softmax é uma variação da regressão logística para predição de múltiplas classes, diferente da regressão logística onde é possível classificar somente duas classes. Essa Regressão é aplicada em problemas onde é preciso trabalhar com mais de duas classes, e cria modelos para cada uma das classes que existem no conjunto de dados.

Exemplos:

A partir de uma imagem de um animal, classificar a qual rótulo um determinado animal pertence. Sendo:

- Rótulo -1, se o animal é um peixe.
- Rótulo 0, se o animal é um gato.
- Rótulo 1, se o animal é um cachorro.

A partir de uma imagem de um raio X, classificar a qual rótulo um determinado paciente pertence. Sendo:

- Rótulo -1, se o paciente possui uma doença.
- Rótulo 0, se o paciente está saudável.
- Rótulo 1, se o paciente possui alguma fratura.

A partir dos exemplos supracitados temos 3 possíveis classes, portanto seria necessária uma regressão Softmax para realizar a tarefa.

Questão 3)

Nas árvores de decisão, temos dois tipos de nós:

- O nó folha que é rotulado como uma função que geralmente é a constante que minimiza a função de custo,
- O nó de divisão que contém um teste condicional baseado nos valores do atributo.

Uma regra de divisão é guiada por uma medida denominada “goodness of split”, que indica quão bem um dado atributo discrimina as classes e ela é usada para selecionar o atributo que maximiza essa medida. Uma regra de divisão, funciona como uma heurística, olhando para a frente um passo, e gera testes nesses passos, e para cada teste possível, o sistema hipoteticamente considera os subconjuntos dos dados obtidos e escolhe o teste que maximiza (ou minimiza) algumas funções heurísticas sobre os subconjuntos. Portanto, a proposta natural é rotular cada subconjunto da divisão por sua classe mais frequente e escolher a divisão que tem menores erros.

Questão 4)

Nem sempre a utilização de apenas um tipo de aprendizado resolve o problema de forma satisfatória. Comumente um modelo tem uma boa precisão para alguns conjuntos de dados e baixa para outros. Dessa forma, uma alternativa é a utilização de vários modelos diferentes, e a extração do resultado final através de cálculos como média aritmética, média ponderada e o rótulo(valor) mais votado.

Exemplo:

Um comerciante quer saber a variação de preço que seus produtos precisarão ter para que ele consiga manter uma certa margem de lucro durante o período de um ano. Portanto, fatores como a variação de preço da matéria prima, variação da mão de obra, quantidade de produtos que serão vendidos, quantidade produtos fabricados com defeito, dentro outros, devem ser previstos.

Sendo assim, um determinado modelo pode não gerar uma solução satisfatória, e, neste caso, seria adequado utilizar vários modelos como os métodos de regressão logística, árvore de decisão, ou até mesmo Random Forest, ou seja, um Aprendizado por Comitê (ensemble learning), para obter boas estimativas.

Questão 5)

a) Overfitting ou sobreajuste dos dados, é o processo em que o modelo adotado funcionou muito bem para um caso específico ou o conjunto de treino, porém, o algoritmo está tão ajustado ao conjunto de dados de testes e/ou treino que não funciona adequadamente para novos dados, ou seja, ocorre uma má generalização. Exemplo: Foi pedido a um programador que fizesse uma análise sobre um banco de dados de medicamentos para descobrir quais os possíveis medicamentos que ajudam a combater a Covid, sendo que esses medicamentos foram tomados pelas pessoas que curaram a doença. Se estiver no banco de dados poucos dados

apontando que pessoas tomaram cloroquina curaram da Covid, e o programador não ajustar corretamente os parâmetros do modelo, então o algoritmo irá generalizar que a cloroquina ajuda a combater a Covid.

b) Uma hipótese é uma função que melhor descreve o objetivo no aprendizado de máquina supervisionado, ou seja, é um modelo candidato que se aproxima de uma função de destino para mapear entradas em saídas e pode ser avaliada e usada para fazer previsões. Exemplo: no algoritmo CART uma hipótese poderia ser a função que determina um ponto de corte de maneira a obter nós mais puros, para que consequentemente seja possível, a partir de uma entrada, prever uma saída (classificação).

c) É como as hipóteses são pesquisadas no espaço de hipóteses, e a preferência de hipóteses sobre outras, ou seja, é a forma como o algoritmo busca a hipótese que melhor se ajusta aos dados de treinamento. O viés de busca define como as hipóteses são pesquisadas no espaço de hipóteses. Por exemplo, o algoritmo ID3, que é utilizado para indução de árvores de decisão, tem como viés de busca a preferência por árvores de decisão com poucos nós.