Prova de Aprendizado de Máquina

MBA em Ciência de Dados

Prof. André Carlos Ponce de Leon F. de Carvalho

**Observações:**

* A interpretação das questões faz parte da prova.
* Dúvidas e omissões, usar sua interpretação e criatividade e definir as condições necessárias.
* A nota final da avaliação será no máximo 10, ainda que o aluno acerte todas as questões integralmente e também a questão bônus.

1. Escreva qual é o viés de busca e o viés de representação dos algoritmos abaixo (2,5 pontos).

Algoritmo de Hunt

Rede neural MLP

1. Escreva 3 diferenças entre os algoritmos Bagging e Boosting (2,5 pontos).
2. Usando no máximo 500 caracteres (incluindo espaços em branco) mencionar 2 deficiências comuns a uma rede neural perceptron e SVM, e qual a solução utilizada para superar essas deficiências em cada uma delas? (2,5 pontos).
3. Por que a acurácia simples não é uma boa medida para tarefas de classificação que utilizam conjuntos de dados desbalanceados? (2,5 pontos).

**Questão Extra**

1. Desenhar como a árvore de decisão abaixo divide o espaço de busca ao lado da árvore (2,5 pontos).

Não

Sim

Sim

Não

Contrata

Contrata

Não contrata

Estudo

Experiência

Não Contrata

Contrata

Não

Sim

Não Contrata

Contrata

Não

Sim

Boa sorte!

**Respostas:**

1. **Escreva qual é o viés de busca e o viés de representação dos algoritmos abaixo (2,5 pontos).**

**Algoritmo de Hunt**

**Rede neural MLP**

O algoritmo de Hunt é um modelo baseado em procura tendo como viés de representação uma árvore de decisão em que cada nó interno é representado por uma pergunta referente ao valor de um atributo e cada nó externo está associado a uma classe.

Usa a estratégia gulosa como viés de busca particionando as características (atributos) de forma hierárquica que é chamado de divisão e conquista, buscando minimizar a impureza de cada nó. O conjunto das regras de decisão usado é bastante simples:

Seja Xt o conjunto de objetos de treinamento que atingem o nó t:

1. Se todos os objetos de Xt ∈ a mesma classe y então o nó t é um nó folha rotulado pela classe y;
2. Se não, selecionar um atributo preditivo teste para dividir Xt;
   * Dividir Xt em subconjuntos usando valores desse atributo;
   * Aplicar o algoritmo a cada subconjunto gerado.

O algoritmo de redes neurais artificiais MLP (multi-layer perceptron) é baseado em técnicas geométricas tendo como viés de representação uma hipótese por um conjunto de valores reais associados aos pesos das conexões da rede.

O viés de busca usa uma ou mais camadas intermediárias de neurônios e uma camada de saída que permite resolver problemas não linearmente separáveis. Usa um conjunto de matrizes de números reais e a Correção de Erro como forma de ajuste, tendo sido originalmente treinada com o algoritmo backpropagation. Na primeira camada, cada neurônio aprende uma função que define um hiperplano que divide o espaço de entrada em duas partes. Cada neurônio da camada seguinte combina um grupo de hiperplanos definidos pelos neurônios a camada anterior, formando regiões convexas. Os neurônios da camada seguinte combinam um subconjunto das regiões convexas em regiões de formato arbitrário. E cada neurônio da camada de saída está associada a uma das classes presentes no conjunto de dados.

1. **Escreva 3 diferenças entre os algoritmos Bagging e Boosting (2,5 pontos).**

**Bagging**

|  |  |
| --- | --- |
| Bagging (Bootstrap Aggregating) | Boosting |
| Comitês que fazem uma combinação paralela de algoritmos | Comitês que fazem uma combinação sequencial de algoritmos |
| Amostra para treinar os modelos é definida usando bootstrap (seleciona um exemplo da base de dados e coloca ele de volta para a próxima seleção), mantendo o número original de observações | Usa todas as observações da base de dados para treinar os modelos |
| Reduz a variância do algoritmo final usando um conjunto de algoritmos com modelos-base que já possuem uma boa variância | Reduz o viés do algoritmo final usando um conjunto de algoritmos com modelos-base fracos (alto viés), atribuindo um peso pelo desempenho de cada modelo para obter um algoritmo global forte |

1. **Usando no máximo 500 caracteres (incluindo espaços em branco) mencionar 2 deficiências comuns a uma rede neural perceptron e SVM, e qual a solução utilizada para superar essas deficiências em cada uma delas? (2,5 pontos).**

Ambos se restringem a problemas linearmente separáveis. O SVM usa Kernels para transformar o espaço de entrada em um espaço de características e a RN Perceptron faz a adição de camadas.

Ambos se limitam a classificações binárias. O SVM trata multiclasses com a adaptação interna do algoritmo e a decomposição do problema em vários problemas binários, sendo essa última a mais comum. Para a RN Perceptron a forma mais simples e frequente é associar um neurônio da camada de saída para cada classe.

1. **Por que a acurácia simples não é uma boa medida para tarefas de classificação que utilizam conjuntos de dados desbalanceados? (2,5 pontos).**

Porque os modelos tendem a privilegiar a classe majoritária.

A acurácia simples mede o % de acertos global e o modelo tente a concentrar seu aprendizado na classe majoritária que é mais numerosa.

Então o modelo pode estar acertando bem a classe majoritária e sendo muito ruim para a classe minoritária que pode ser o objeto de interesse do modelo como por exemplo em detecção de fraudes, de doenças etc e essa métrica não fará essa identificação.

Para a classificação de dados desbalanceados, a sensibilidade pode ser mais interessante que a especificidade.

Elas podem ser combinadas em uma medida simples, que busca atender as duas demandas. Dois exemplos são as médias geométricas e a acuraria balanceada.

1. **Desenhar como a árvore de decisão abaixo divide o espaço de busca ao lado da árvore (2,5 pontos).**

Não

Sim

Sim

Não

Contrata

Contrata

Não contrata

Estudo

Experiência

Não Contrata

Contrata

Não

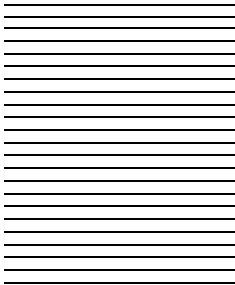
Sim

Não Contrata

Contrata

Não

Sim



10

10

20

5