Prova de Aprendizado de Máquina

MBA em Ciência de Dados

Prof. André Carlos Ponce de Leon F. de Carvalho

***Aluno: Bruno Birolli Nasser***

**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

**Resposta:** Algoritmo de Hunt tem representação baseado em arvores de decisão, nos internos e raiz são atributos de predição e nós externos (folhas) atributo alvo.

Rede neural MLP é um sistema distribuído inspirado no cérebro humano. Os neurônios são representados por unidades de processamento que aplicam uma função de ativação sobre as entradas. Essas redes podem conter uma ou mais camadas intermediárias de neurônios.

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

**Resposta:**

Bagging: Combinação paralela, induz classificadores, cada classificador é induzido por uma amostra diferente do conjunto de treinamento

Boosting: Combinação sequencial, induz modelos preditivos, classificador base é uma árvore de decisão Stump

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).

**Resposta:** Apenas lidam com problemas linearmente separáveis. Ajuste de pesos não leva em conta a distância entre saída e resposta desejada apenas se elas são iguais ou diferentes.

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

**Resposta:** Acurácia simples leva em conta em seu calculo tanto os verdadeiros positivos quanto os verdadeiros negativos, ou seja, o tanto que se acerta do objetivo e do que não se pretende prever. Se a base de dados está desbalanceada a acurácia pode indicar um valor alto e ao mesmo tempo não significar que o modelo está com um bom desempenho já que este pode estar acertando muito bem os verdadeiros negativos e não necessariamente o que se deseja prever. Neste caso a sensibilidade é mais interessante para analisar o desempenho do modelo.

**Questão Extra**

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

(\*) Para resolver o exercício troquei as posições entre o Sim e o Não originais senão ficava todos os campos com “contrata” e não fazia sentido.

Não

Sim

**Sim\***

**Não\***

Contrata

Contrata

Não Contrata

Não contrata

Estudo

Experiência

Não Contrata

Contrata

Não

Sim

Não Contrata

Contrata

Não

Sim

10

20

5

10

Contrata

Contrata

Contrata

Não contrata

Boa sorte!