

Universidade de São Paulo

Escola de Artes, Ciências e Humanidades

Disciplina: SIN 5016 – Aprendizado de Máquina

Docente: Prof. Dr. Clodoaldo A Moraes Lima

Discentes: _____ No. USP: _____

Lista de Exercícios

1ª Questão) Com relação a redes neurais artificiais, responda os itens abaixo

a) Sabendo que todo raciocínio está fundamentado em três modalidades de inferência: dedução, indução e abdução. A dedução parte de uma regra geral e chega em uma regra particular. Já a indução parte de regras particulares e chega em regras gerais, gerando assim conhecimento novo.

Procure então fornecer argumentos que sustentem as seguintes afirmações:

a.1) o princípio de operação do método do gradiente em otimização é equivalente à dedução.

a.2) o princípio de operação de redes neurais que foram sujeitas a treinamento supervisionado é equivalente à indução

b) Apresente um pseudo-código para o algoritmo Back-Propagation, considerando a correção do erro em lote e sequencial. O aluno deve apresentar as equações de cálculo da saída da rede, do gradiente e atualização dos pesos.

c) Descreva quais parâmetros o usuário deve definir para executar uma Rede Neural MLP. Apresente uma estratégia para definição de cada um desses parâmetros.

d) Apresente duas vantagens e duas desvantagens dos algoritmos de segunda e primeira ordem para atualização dos pesos.

e) Alguns autores sugerem inserir um termo de penalidade na função objetiva para controlar a suavidade do mapeamento produzido pela RNA. Descreva detalhadamente esta abordagem e apresente a função objetiva a ser minimizada. Qual o significado do parâmetro de regularização?

f) Faça um estudo comparativo entre Redes Neurais Multilayer Perceptron e Redes Neurais de Funções de Base Radial. Apresente **pelo menos duas** semelhanças e duas diferenças entre RBF e MLP. Além disso, discuta sobre os parâmetros a serem determinados em cada arquitetura.

g) Explique o que é capacidade de generalização em treinamento supervisionado de redes neurais artificiais e como a disponibilidade de um conjunto de treinamento e de um conjunto de validação pode ser empregada visando maximizar esta capacidade.

h) Explique o que é maldição da dimensionalidade. Como a maldição da dimensionalidade afeta o desempenho de um modelo neural.

i) Qual a função do termo momento na função objetiva.

i) Na síntese de um classificador de padrões empregando uma rede neural, a qual é treinada a partir de dados de entrada-saída disponíveis, percebeu-se que o desempenho do classificador resultante era insatisfatório em aplicações práticas. Apresente possíveis razões para este insucesso sabendo que:

- i.1) não há nenhuma informação disponível sobre o processo de treinamento do classificador;
- i.2) sabe-se que foi empregado um conjunto de treinamento, mas não foi adotado um conjunto de validação;
- i.3) sabe-se que foi empregado tanto conjunto de treinamento como de validação.

2ª Questão) Considerando uma Rede Neural Artificial (RNA), pede-se

- a) Apresente um pseudocódigo para o algoritmo BackPropagation, considerando a correção do erro em lote e sequencial. O aluno deve apresentar as equações de cálculo da saída da rede, do gradiente e atualização dos pesos.
- b) Descreva quais parâmetros o usuário deve definir para executar uma Rede Neural. Apresente uma estratégia para definição de cada um desses parâmetros.
- c) Apresente duas vantagens e duas desvantagens dos algoritmos de segunda e primeira ordem para atualização dos pesos.
- d) Em problemas de predição um passo à frente, o processo de treinamento pode trabalhar com partições sequenciais dos dados da série, ou então empregar uma partição que será denominada aqui de aleatória. Explique qual a melhor forma para trabalhar com uma Rede Neural estática e uma Rede Neural Recorrente.
- e) Na tabela abaixo, parte do algoritmo para treinamento de um *Multilayer Perceptron* com BackPropagation do erro é apresentada. Considerando essa lógica de treinamento, explique a função das variáveis α , δ_k , δ_{in_j} e δ_j , dando destaque para o relacionamento **entre essas variáveis e entre essas variáveis e aquelas que representam os pesos** da rede neural.

Passo 3: Cada unidade de entrada ($X_i, i = 1 \dots n$) recebe um sinal de entrada x_i e o dissipa para todas as unidades na camada acima (unidades escondidas).

Passo 4: Cada unidade escondida ($Z_j, j = 1 \dots p$) soma suas entradas pesadas,

$$z_{in_j} = v_{0j} + \sum_{i=1}^n x_i v_{ij}$$

aplica sua função de ativação para computar seu sinal de saída,

$$z_j = f(z_{in_j})$$

e envia o sinal para todas as unidades na camada acima (unidades de saída).

Passo 5: Cada unidade de saída ($Y_k, k = 1 \dots m$) soma suas entradas pesadas

$$y_{in_k} = w_{0k} + \sum_{j=1}^p z_j w_{jk}$$

e aplica sua função de ativação para computar seu sinal de saída,

$$y_k = f(y_{in_k})$$

| | |
|---|--|
| <p>Passo 6: Cada unidade de saída ($Y_k, k = 1 \dots m$) recebe uma classificação correspondente ao padrão de entrada, computa seu termo de erro de informação</p> $\delta_k = (t_k - y_k) f'(y_{in_k})$ <p>calcula seu termo de correção de pesos</p> $\Delta w_{jk} = \alpha \delta_k z_j$ <p>calcula seu termo de correção de bias</p> $\Delta w_{0k} = \alpha \delta_k$ <p>e envia δ_k para as unidades cada camada abaixo.</p> | <p>Passo 7: Cada unidade de saída ($Z_j, j = 1 \dots p$) soma suas entradas delta (vindas das unidades da camada acima)</p> $\delta_{in_j} = \sum_{k=1}^m \delta_k w_{jk}$ <p>multiplica pela derivada de sua função de ativação para calcular seu termos de erro de informação</p> $\delta_j = \delta_{in_j} f'(z_{in_j})$ <p>calcula seu termo de correção de pesos</p> $\Delta v_{ij} = \alpha \delta_j x_i$ <p>e calcula seu termo de correção de bias</p> $\Delta v_{0j} = \alpha \delta_j$ |
|---|--|

- f) Discuta a função da camada escondida de uma rede neural artificial multicamadas. Forneça um exemplo didático para ilustrar sua discussão. Inclua em sua discussão uma relação entre o papel de uma camada escondida e problemas não linearmente separáveis.
- g) Com relação aos algoritmos de otimização irrestrita, cite dois algoritmos de primeira e segunda ordem e apresente uma vantagem e uma desvantagem de cada um.

3ª Questão) No contexto de redução de dimensão e visualização de dados, análise de componentes principais (PCA, do inglês principal component analysis) é geralmente adotada. Basicamente, o PCA define um hiperplano de projeção de dimensão menor ou igual à dimensão do espaço original dos dados. O critério de otimização do PCA é minimizar o somatório da distância dos dados originais ao hiperplano. Como consequência da aplicação deste critério, o posicionamento ótimo do hiperplano de projeção faz com que os dados projetados apresentem o máximo espalhamento possível, de modo que os eixos do hiperplano representam as direções de maior variância dos dados. Abusando um pouco da terminologia, pode-se denominar de análise de componentes principais não-lineares (NLPCA, do inglês nonlinear principal component analysis) a toda iniciativa equivalente ao PCA, mas que toma uma hipersuperfície em lugar de um hiperplano de projeção. Como esta hipersuperfície pode ser produzida por uma rede neural MLP, pode-se empregar a técnica de NLPCA para um problema de classificação de padrões. A ideia é usar uma MLP para o papel de NLPCA e uma outra MLP para o classificador, e treinar ambas as MLPs simultaneamente. Explique detalhadamente qual deve ser a configuração de uma rede neural para que esta possa ser empregada como NLPCA.

4ª Questão) Considere uma Rede Neural Convolutiva (CNN) com uma camada de convolução com strider maior que um, uma camada de pooling, uma camada totalmente com função de ativação relu e uma camada de saída com função de ativação softmax. A função custo é dada por: $J(\theta) = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^{ns} d_k(n) \ln y_k(n)$.

$d_k(n)$ - k-ésima saída desejada

$y_k(n)$ - k-ésima saída da rede neural

- a) Apresente uma motivação biológica para estudo de CNN

b) Apresente a formula de atualização dos filtros da CNN.

5ª Questão) Com relação as Redes Neurais Convolucionais (CNNs), responda os itens abaixo.

- Apresente duas motivações para o emprego de Redes Neurais Convolucionais.
- Explique o que são pesos compartilhados e apresente uma vantagem para o emprego destes.
- Explique detalhadamente qual a função da camada de convolução e pooling.
- Explique detalhadamente como funciona o dropout.
- Descreva como é realizado o treinamento de uma CNN
- Explique como podemos utilizar uma CNN treinada para extração de característica.
- Explique o que seria os autoencoders.
- Explique como funciona autoencoders variacional e robustos.
- Dado uma CNN pretreinada para reconhecimento facial. Como podemos utilizá-la para extração de características faciais?
- Como podemos utilizar os autoencoders para atacar o problema da maldição da dimensionalidade?

6ª Questão) A tabela abaixo apresenta os exemplos de decisão sobre que ações um robô deve tomar enquanto dirige um carro.

| Semáforo | Luz de Freio | Distância do carro frente (metros) | Ação |
|----------|--------------|------------------------------------|------------------|
| Verde | Não | 10 | Andar |
| Verde | Sim | 50 | Andar |
| Verde | Sim | 58 | Parar |
| Vermelho | Sim | 60 | Parar |
| Vermelho | Não | 70 | Parar |
| Amarelo | Não | 80 | Virar à Esquerda |
| Amarelo | Sim | 30 | Virar à Direita |
| Amarelo | Não | 10 | Virar à Esquerda |

Suponha uma MLP com 3 neurônios na camada escondida com função de ativação softmax e saída com função de ativação softmax para modelar este problema. Considere uma taxa de aprendizado igual a 0.5. Pede-se

- Apresente a codificação e o préprocessamento adotado.
- Inicialize todos pesos. Calcule a saída da rede neural e monte a matriz de confusão
- Mostre detalhadamente os cálculos que o algoritmo backpropagation executará na primeira época.
- Monte a curva roc após treinar a rede neural uma época.

7ª Questão) Considerando Árvore de Decisão, responda os itens abaixo.

- Apresente duas motivações para utilizar uma Árvore de Decisão
- Explique como é realizado o processo de indução de uma árvore de decisão: algoritmo de indução, medidas de escolha de nós, parâmetros livres (discuta), estratégias de pré-poda e pós-poda.
- Randon Forests: qual a motivação, estratégias para construção das árvores e para combinação de resultados, parâmetros livres.
- Explique como podemos utilizar uma árvore de decisão para seleção de atributos

- e) Dado um atributo numérico, explique detalhadamente como devemos proceder para utilizá-lo na construção de uma árvore de decisão.

8ª Questão) Considere o problema de aprender o conceito sobre pacientes doentes (+) ou saudáveis (-). Foram coletados os seguintes exemplos

| Conjunto de treinamento | | | |
|-------------------------|---------|---------|--------|
| Idade | Exame A | Exame B | Classe |
| 20 | H | C | + |
| 25 | R | B | + |
| 28 | R | C | + |
| 23 | H | B | + |
| 19 | J | B | + |
| 35 | R | C | - |
| 30 | J | C | - |
| 38 | J | B | - |
| 40 | H | C | + |
| 24 | J | C | - |
| 29 | R | C | - |
| 20 | J | B | + |
| 19 | R | C | - |

| Conjunto de Teste | | | |
|-------------------|---|---|---|
| 20 | R | B | + |
| 22 | J | B | - |
| 30 | J | C | - |
| 31 | R | C | + |
| 29 | H | C | + |

- a) Utilizando o algoritmo de indução de árvores de decisão, construa a árvore correspondente (sem poda e sem número mínimo de exemplos em cada folha), utilizando o critério de ganho de informação para selecionar atributos para este conjunto de exemplos. Anote, para cada nível da árvore de decisão, o valor do ganho de informação calculado para cada atributo, bem como aquele escolhido para particionar os exemplos. Se houver empate entre valores do ganho de informação, escolha o primeiro (na ordem da tabela acima). Se for necessário realizar a discretização dos atributos, descreva detalhadamente o processo adotado.
- b) Análise de desempenho: use a árvore de decisão produzida em (a) para classificar os exemplos do conjunto de teste. Informe a precisão e matriz de confusão da árvore para esses exemplos. Discuta resumidamente os resultados.
- c) Aplique dois critérios de poda visto em sala de aula a árvore gerada no item b). Em seguida apresente a matriz de confusão para o conjunto de teste.
- d) Aplique dois critérios de poda visto em sala de aula a árvore gerada no item b). Em seguida apresente a matriz de confusão para o conjunto de teste.

9ª Questão) Considerando Máquinas de Vetores Suporte, responda os itens abaixo.

- e) Descreva detalhadamente o SVM e suas variações (pelo menos 2). Apresente a formulação do primal e do dual.
- f) Descreva detalhadamente como SVM pode ser empregado para problemas com múltiplas classes. Apresente três estratégias e explique detalhadamente cada uma.
- g) Explique detalhadamente os três métodos alternativos para treinamento de SVM (Chunking, Algoritmo de Osuna, SMO).

- h) Explique o que significa o Truque do Kernel (Kernel Trick).
- i) Explique detalhadamente como podemos aplicar dimensão VC para selecionar o tipo de kernel.

10ª Questão) Considerando uma Rede Neural Artificial (RNA), pede-se

- a) Desenhe uma RNA, com número mínimo de neurônios na camada escondida capaz de modelar o problema do Ou-Exclusivo.
- b) Considerando o item c), inicialize todos os pesos iguais 0.1 e calcule a saída da RNA. Em seguida, mostre os cálculos que o algoritmo Levenberg–Marquardt de atualização dos pesos executará na primeira época para apenas uma instância
- c) Apresente duas vantagens e duas desvantagens dos algoritmos de segunda e primeira ordem para atualização dos pesos
- d) Alguns autores sugerem inserir um termo de penalidade na função objetiva para controlar a suavidade do mapeamento produzido pela RNA. Descreva detalhadamente esta abordagem e apresente a função objetiva a ser minimizada. Qual o significado do parâmetro λ ?

11ª Questão) Considerando Comitê de Máquinas

- a) Faça um comparativo entre a abordagem proposta por Rosen (Descorrelated Neural Networks) e por Liu (Learning via Negative Correlation)
- b) Explique detalhadamente as 3 fases de construção do ensemble. Para cada fase apresente 2 estratégias
- c) Apresente três vantagens e duas desvantagens do Ensemble e da Mistura de Especialista.
- d) Explique detalhadamente o papel gating em Mistura de Especialista. Apresente estratégias de inicialização para gating e especialistas baseada em aprendizado não supervisionado.

12ª Questão) Considerando Máquinas de Vetores Suporte

- a) Faça uma breve discussão sobre as variações (pelo menos 3) de SVM. Apresente a formulação primal e dual de cada abordagem.
- b) Descreva detalhadamente como SVM pode ser empregado para problemas com múltiplas classes.
- c) Explique o que significa o Truque do Kernel (Kernel Trick).
- d) Explique detalhadamente os três métodos alternativos para treinamento de SVM (Chunking, Algoritmo de Osuna, SMO).
- e) Descreva detalhadamente Support Vector Clustering (SVC). Qual o problema dessa abordagem. Explique detalhadamente como podemos aplicar dimensão VC e SVC para selecionar o tipo de kernel.
- f) Explique detalhadamente Extreme Support Vector Machine

12ª Questão) Apresente uma arquitetura de uma MLP que pode ser empregada para modelar componentes principais. Em seguida apresente todos passos e as derivadas para treinamento da rede neural especificada e o cálculo das componentes.

13ª Questão) Suponha que uma MLP com 3 neurônios na camada escondida taxa de aprendizado seja 0.5, seja utilizada para modelar o problema do Ou-Exclusivo. Utilize o menor número de saídas possível para o problema.

- Inicialize todos pesos com 0.1. Calcule a saída da rede neural e monte a matriz de confusão
- Mostre os cálculos que o algoritmo backpropagation executará na primeira época para o peso a_{22} da camada de entrada e b_{11} para camada intermediária, considerando apenas a primeira instância

14ª Questão) Discuta a função da camada escondida de uma rede neural artificial multicamadas. Forneça um exemplo didático para ilustrar sua discussão. Inclua em sua discussão uma relação entre o papel de uma camada escondida e problemas não linearmente separáveis.

15ª Questão) Considere os conjuntos de dados descritos nos quadros abaixo:

i)

| | | | |
|---|-----------------------------|-----------------------------|-----|
| Seeds Data Set (adaptado de UCI – Machine Learning Repository / doado pela John Paul II Catholic University of Lublin e Cracow University of Technology) | | | |
| Características do conjunto de dados: | Multivariado | Número de instâncias | 210 |
| Características dos atributos: | Numérico / categórico | Número de atributos | 7 |
| Tarefas associadas: | Classificação / agrupamento | Valores faltantes | Não |

No conjunto de dados estão armazenadas as seguintes informações sobre sementes:

1. área (A): numérico
2. perímetro (P): numérico
3. compacidade ($C = 4 \cdot \pi \cdot A / P^2$): numérico
4. comprimento do núcleo: numérico
5. largura do núcleo: numérico
6. coeficiente de assimetria: numérico
7. comprimento do núcleo do sulco: numérico
8. variedade: kama, rosa, canadense (atributo classe)

ii)

| | | | |
|--|--------------|-----------------------------|-------|
| Bank Marketing Data Set | | | |
| Resumo: Os dados estão relacionados com campanhas de marketing direto (via chamadas telefônicas) de uma instituição financeira portuguesa. A meta do problema de classificação é decidir se o cliente aceitará o produto. Frequentemente, é necessária a realização de mais de um contato com o mesmo cliente, a fim de verificar se o produto seria aceito ou não. | | | |
| Data Set Characteristics: | Multivariate | Number of Instances: | 45211 |

| | | | |
|----------------------------|----------------------------|-----------------------|-----|
| Attribute Characteristics: | Categorical, Integer, Real | Number of Attributes: | 17 |
| Associated Tasks: | Classification | Missing Values? | N/A |

Fonte: [Moro et al., 2011] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 2011. EUROSIS (<http://hdl.handle.net/1822/14838>)

Informações sobre o conjunto de dados:

Existem dois conjuntos de dados:

- 1) bank-full.csv com todos os dados (instâncias) disponíveis, ordenados por data (de maio de 2009 a novembro de 2010)
- 2) bank.csv com 10% dos dados (4521 instâncias), randomicamente selecionados de bank-full.csv.

Informações sobre os atributos:

(dados sobre os clientes)

- Idade (numérico)
- Tipo de emprego (categórico: “admin”, “desconhecido”, “desempregado”, “gerente”, “dona de casa”, “empresário”, “estudante”, “operário”, “autônomo”, “aposentado”, “técnico”, “prestador de serviço”)
- Estado civil (categórico: “casado”, “divorciado”, “solteiro”)
- Grau de instrução (categórico: “desconhecido”, “primário”, “secundário”, “terciário”)
- Possui crédito? (binário: sim / não)
- Renda anual (numérico)
- Possui hipoteca imobiliária? (binário: sim / não)
- Possui empréstimo? (binário: sim / não)

(dados sobre o último contato da campanha atual)

- Tipo de comunicação do contato (categórico: “desconhecido”, “telefone”, “celular”)
- Dia do mês do contato (numérico)
- Duração do contato em segundos (numérico)

(outros atributos)

- Número de contatos executados durante a campanha para este cliente (numérico)
- Número de dias que se passaram desde o último contato com o cliente considerando campanha anterior (numérico / -1 significa que o cliente não foi contatado antes)
- Número de contatos executados com este cliente antes desta campanha (numérico)
- Resultado da campanha anterior com relação ao cliente (categórico: “desconhecido”, “outros”, “falha”, “sucesso”)

(variável de classe)

- O cliente contatado aceitou o produto? (binário: sim (5288 instâncias) / não (39923 instâncias))

a) Para resolver os problemas de classificação descritos nos quadros i) e ii), você considera que seria mais adequado usar uma rede neural artificial (com aprendizado supervisionado) ou uma árvore de decisão? Justifique sua resposta. Em sua justificativa, considere a teoria que você estudou sobre redes neurais artificiais e árvores de decisão, e argumente em favor de sua escolha usando informações presentes na descrição do conjunto de dados.

b) Considerando que alguns atributos descritivos dos problemas acima não possuem um poder de expressividade muito bom, e que há algumas classes mal representadas e difíceis de aprender, que tipo de alteração você poderia fazer para atender melhor ao problema? Por que essa alteração vai ajudar na construção de um modelo de classificação melhor?

16ª Questão) Assinale verdadeiro (V) ou falso (F). Lembre-se que um item assinalado incorretamente anula um item corretamente

(.....) A escolha adequada da taxa de aprendizado em Redes Neurais Artificiais é muito importante para assegurar a estabilidade da convergência do processo de aprendizado iterativo, pois taxas pequenas permitem um aprendizado mais lento porém mais consistente, mas com o perigo de cair em mínimos locais, enquanto taxas maiores permitem um aprendizado mais rápido a custo muitas vezes de desestabilização (oscilação)

(.....) A ativação de um neurônio na rede MLP se dá pelo produto interno entre seu vetor de pesos e o vetor de entradas, seguida pela aplicação da função de ativação, geralmente do tipo sigmoideal. Por outro lado, a ativação de um neurônio na rede RBF se dá pelo cálculo da distância euclidiana entre o vetor de pesos do neurônio e o vetor de entradas. Quanto mais distante o vetor de entrada do vetor de pesos, menor a ativação do neurônio.

(.....) as redes MLP tendem a se dar melhor no caso de número elevado de entradas, quando comparado às redes RBF. Conforme aumenta o número de entradas, o número de funções de base radial tende a crescer exponencialmente, caso se queira manter o mesmo nível de desempenho. Essa lei é conhecida como “maldição da dimensionalidade”.

(.....) motivado pelas decisões de projeto, o treinamento de uma rede neural RBF se dá em um único passo de cálculo (usando uma fórmula matemática fechada), representado pela pseudo-inversão de uma matriz e por produtos entre matrizes e vetores. Por outro lado, o treinamento de uma rede neural MLP envolve aplicações iterativas de um processo de ajuste incremental do vetor de pesos, sendo necessário definir a cada iteração um passo e uma direção de ajuste.

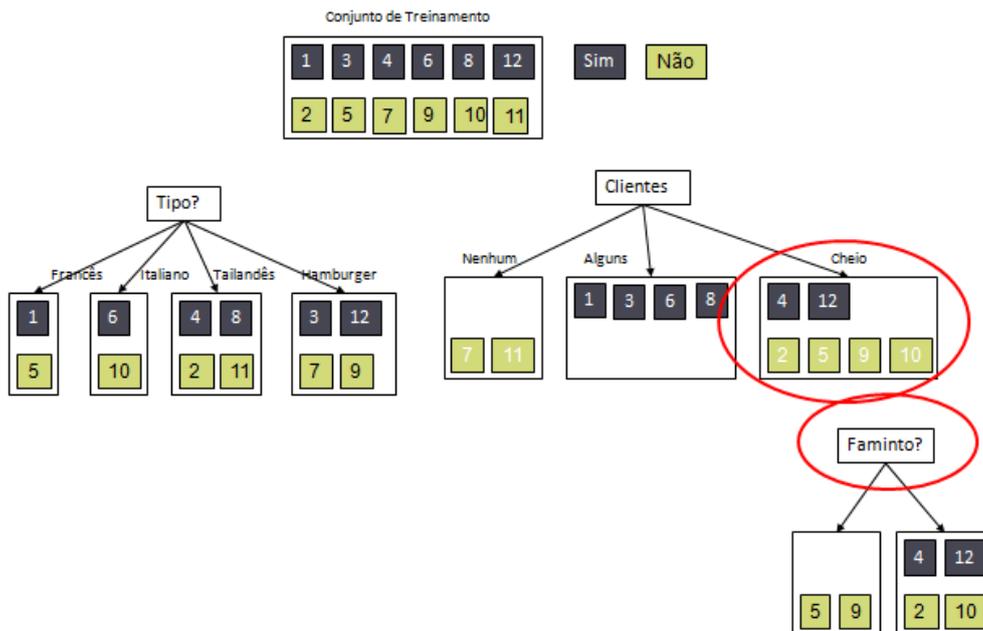
(.....) os únicos parâmetros ajustáveis na rede RBF são os pesos da camada de saída, enquanto que na rede MLP os pesos da camada intermediária são também ajustáveis. Com isso, o projetista da rede RBF deve definir o número de neurônios, o centro das funções de base radial e parâmetros de dispersão desta função de base radial.

() Holdout reserva uma certa quantidade de dados para treinamento e o restante para teste (podendo ainda usar parte para validação). Comumente esta estratégia usa 1/3 dos dados para teste e o restante para treinamento, escolhido randomicamente. Para conjunto de dados randômicos é interessante assegurar que a amostragem randômica seja feita de tal maneira que garanta que cada classe é apropriadamente representada tanto no conjunto de treinamento quanto no conjunto de teste. Este procedimento é chamado de **holdout repetitivo**. Se for realizado apenas uma divisão do conjunto de dados, a estimativa da taxa de erro vai ser enganosa se acontecer de termos uma divisão ruim. Visando amenizar tendências, emprega-se **holdout estratificado**, o qual consiste em repetir-se todo o processo de treino e teste várias vezes com diferentes amostragens randômicas.

() O bootstrap é um procedimento estatístico de amostragem com reposição. Considerando um conjunto de dados com n instâncias, n instâncias são escolhidas aleatoriamente. Uma instância não é retirada do conjunto de dados original quando ela é escolhida para compor o conjunto de treinamento, ou seja, a mesma instância pode ser selecionada várias vezes durante o procedimento de amostragem.

() No n fold cross validation o conjunto de dados é dividido em n partições de tamanhos aproximadamente iguais e, de maneira rotativa, cada uma delas é usada para teste enquanto as restantes são usadas para treinamento. Este procedimento é repetido n vezes. Para conjunto de dados pequeno, geralmente n é escolhido igual ao número de instâncias no conjunto de dados. Este é também conhecido como leave one out.

16ª Questão) Analise o gráfico abaixo e responda as questões na sequência:



- Entre os atributos TIPO e CLIENTES, qual tem maior entropia? Qual deles deve ser escolhido para iniciar a construção de uma árvore? Apresente os cálculos que justifique sua resposta.
- Se o atributo CLIENTES é usado na árvore, o que ocorre com os nós no fim dos ramos “NENHUM” e “ALGUNS”?
- Se pararmos o processo de construção da árvore após a inserção dos nós com os atributos CLIENTE e FAMINTO, o que podemos dizer sobre o desempenho de classificação desta árvore?
- Considerando os três atributos (tipo, clientes e faminto), qual seria a árvore de decisão final (apresente todos os cálculos)? Rotule todos os nós folhas, em caso de empate escolha a classe final como SIM. Mostre a matriz de confusão

17ª Questão) Considere o seguinte conjunto de exemplos de treinamento, onde a variável GripeA é a classe:

| Tosse | Febre | Viagem | GripeA |
|-------|-------|--------|--------|
| V | F | F | F |
| F | F | V | F |
| V | V | F | F |
| V | V | V | V |
| F | V | F | F |
| F | F | F | F |
| V | F | V | V |
| F | V | V | V |

- Utilizando o algoritmo de indução de árvores de decisão, construa a árvore correspondente (sem poda e sem número mínimo de exemplos em cada folha), utilizando o critério de ganho de informação para selecionar atributos para este conjunto de exemplos. Anote, para cada nível da árvore de decisão, o valor do ganho de informação calculado para cada atributo, bem como aquele escolhido para particionar os exemplos. Se houver empate entre valores do ganho de informação, escolha o primeiro (na ordem da tabela acima).
- Por que uma árvore de decisão podada que não se ajusta perfeitamente aos dados de treinamento pode ser melhor do que uma árvore não podada?

18ª Questão) Considere os conjuntos de dados descritos nos quadros abaixo:

i)

Seeds Data Set (adaptado de UCI – Machine Learning Repository / doado pela John Paul II Catholic University of Lublin e Cracow University of Technology)

| | | | |
|--|-----------------------------|-----------------------------|-----|
| Características do conjunto de dados: | Multivariado | Número de instâncias | 210 |
| Características dos atributos: | Numérico / categórico | Número de atributos | 7 |
| Tarefas associadas: | Classificação / agrupamento | Valores faltantes | Não |

No conjunto de dados estão armazenadas as seguintes informações sobre sementes:

1. área (A): numérico
2. perímetro (P): numérico
3. compactidade ($C = 4 \cdot \pi \cdot A / P^2$): numérico
4. comprimento do núcleo: numérico
5. largura do núcleo: numérico
6. coeficiente de assimetria: numérico
7. comprimento do núcleo do sulco: numérico
8. variedade: kama, rosa, canadense (atributo classe)

ii)

Bank Marketing Data Set

Resumo: Os dados estão relacionados com campanhas de marketing direto (via chamadas telefônicas) de uma instituição financeira portuguesa. A meta do problema de classificação é decidir se o cliente aceitará o produto. Frequentemente, é necessária a realização de mais de um contato com o mesmo cliente, a fim de verificar se o produto seria aceito ou não.

| | | | |
|----------------------------|----------------------------|-----------------------|-------|
| Data Set Characteristics: | Multivariate | Number of Instances: | 45211 |
| Attribute Characteristics: | Categorical, Integer, Real | Number of Attributes: | 17 |
| Associated Tasks: | Classification | Missing Values? | N/A |

Fonte: [Moro et al., 2011] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 2011. EUROSIS (<http://hdl.handle.net/1822/14838>)

Informações sobre o conjunto de dados:

Existem dois conjuntos de dados:

- 1) bank-full.csv com todos os dados (instâncias) disponíveis, ordenados por data (de maio de 2009 a novembro de 2010)
- 2) bank.csv com 10% dos dados (4521 instâncias), randomicamente selecionados de bank-full.csv.

Informações sobre os atributos:

(dados sobre os clientes)

- Idade (numérico)
- Tipo de emprego (categórico: “admin”, “desconhecido”, “desempregado”, “gerente”, “dona de casa”, “empresário”, “estudante”, “operário”, “autônomo”, “aposentado”, “técnico”, “prestador de serviço”)
- Estado civil (categórico: “casado”, “divorciado”, “solteiro”)
- Grau de instrução (categórico: “desconhecido”, “primário”, “secundário”, “terciário”)

- Possui crédito? (binário: sim / não)
- Renda anual (numérico)
- Possui hipoteca imobiliária? (binário: sim / não)
- Possui empréstimo? (binário: sim / não)

(dados sobre o último contato da campanha atual)

- Tipo de comunicação do contato (categórico: “desconhecido”, “telefone”, “celular”)
- Dia do mês do contato (numérico)
- Duração do contato em segundos (numérico)

(outros atributos)

- Número de contatos executados durante a campanha para este cliente (numérico)
- Número de dias que se passaram desde o último contato com o cliente considerando campanha anterior (numérico / -1 significa que o cliente não foi contatado antes)
- Número de contatos executados com este cliente antes desta campanha (numérico)
- Resultado da campanha anterior com relação ao cliente (categórico: “desconhecido”, “outros”, “falha”, “sucesso”)

(variável de classe)

- O cliente contatado aceitou o produto? (binário: sim (5288 instâncias) / não (39923 instâncias))

a) Para resolver os problemas de classificação descritos nos quadros i) e ii), você considera que seria mais adequado usar uma rede neural artificial (com aprendizado supervisionado) ou uma árvore de decisão? Justifique sua resposta. Em sua justificativa, considere a teoria que você estudou sobre redes neurais artificiais e árvores de decisão, e argumente em favor de sua escolha usando informações presentes na descrição do conjunto de dados.

b) Considerando que alguns atributos descritivos dos problemas acima não possuem um poder de expressividade muito bom, e que há algumas classes mal representadas e difíceis de aprender, que tipo de alteração você poderia fazer para atender melhor ao problema? Por que essa alteração vai ajudar na construção de um modelo de classificação melhor?