

CMC-15 Inteligência Artificial – Terceiro Laboratório –

Professor: Paulo Marcelo Tasinaffo.

Data de Divulgação: primeira semana do segundo bimestre.

Data de Entrega: *até sexta-feira da oitava semana do segundo bimestre*. O atraso na entrega da lista acarretará no desconto de 20% na nota da mesma. Depois da primeira semana de exames a lista de laboratório não será mais aceita pelo professor.

Regulamentos:

1. Esta Lista de Laboratório (LL1) pode ser resolvida em grupos de dois ou três alunos em cada grupo.
2. A média aritmética das notas das Listas Teóricas (LT1) e (LT2) com a ***nota da Lista de Laboratório*** (LL1) comporá a nota final do segundo bimestre da disciplina CMC-15.
3. O ***segundo laboratório*** (LL2) comporá a nota de Exame Final da parte do Prof. Tasinaffo.
4. Todas as listas e laboratórios deverão ser **entregues até a sexta-feira da última semana de aula do Segundo Bimestre (29/11/2024, Sexta-Feira)**. Deverão ser entregues quatro arquivos PDF (por grupo) e separados; quais sejam, LT1.pdf, LT2.pdf, LL1.pdf e LL2.pdf. Estes quatro arquivos deverão ser enviados e entregues via ambiente Classroom ou para o e-mail do Prof. Tasinaffo.
5. O e-mail do Professor Tasinaffo, para eventuais dúvidas, é tasinaffo@ita.br.
6. Serão disponibilizados três temas de laboratório. Entretanto, cada grupo poderá escolher dois temas, entre os três propostos. A escolha é livre.

Tema do Projeto: resolver o problema de *classificação entre democratas e republicanos* utilizando o conceito da pseudo-inversa para treinar uma rede neural com arquitetura RBF.

Atributos: 1. Bebês Deficientes, 2. Divisão de Custos de Projetos de Água, 3. Adoção de Resolução Orçamentária, 4. Congelamento de Taxas Médicas, 5. Grupos Religiosos nas Escolas, 6. Proibição de Testes Anti-Satélite, 7. Mísseis mx, 8. Imigração, 9. Redução de Corporações de Combustíveis Sintéticos, 10. Gastos com Educação, 11. Direito de Superfundo de Processar, 12. Crime, 13. Exportações Isentas de Impostos, 14. Ato de Administração de Exportação da África do Sul, 15. Alvo (Democrata ou Republicano).

Por exemplo, o atributo “10. Gasto com Educação” será “1” se o político for favorável a esta política de investimento na educação, caso contrário será igual a “0”. Leia atentamente das Figuras 01 e 02. A Figura 01 refere-se à capacidade de representação das redes RBF. A Figura 02 refere-se às fórmulas para cálculo dos pesos, respectivamente, da Rede de Regularização e da Rede RBF. A Figura 03 refere-se ao cálculo da Rede de Regularização considerando a teoria de Regularização de Tikonov.

1. Tarefas a Realizar

1.1. Classificador baseado em Redes Neurais com Arquitetura RBF

Utilizando a base de dados fornecida, criar um classificador baseado em Redes RBF que classifique uma entrada (classifique um indivíduo) como republicano (1) ou democrata (0), com base nas informações disponíveis nas outras colunas. Separe 80% das linhas para treinamento e as demais para teste. Descreva este processamento dos dados para prepará-los para o algoritmo de decisão. No final dizer quais foram as porcentagens de acerto nos padrões de treinamento e nos padrões de teste.

Capacidade de Separação de uma Superfície

Temos um corolário para o teorema de Cover na forma de um resultado assintótico célebre que pode ser formulado como:

O número máximo esperado de padrões (vetores) atribuídos aleatoriamente que são linearmente separáveis em um espaço de dimensão m_1 é igual a $2m_1$.

Este resultado sugere que $2m_1$ é uma definição natural para a *capacidade de separação* de uma família de superfícies de decisão tendo m_1 graus de liberdade. De uma certa forma, a capacidade de separação de uma superfície está intimamente relacionada com a noção de dimensão V-C.

Na teoria de Vapnik–Chervonenkis, a dimensão Vapnik–Chervonenkis (VC) é uma medida do tamanho (capacidade, complexidade, poder expressivo, riqueza ou flexibilidade) de uma classe de conjuntos. A noção pode ser estendida para classes de funções binárias. Ela é definida como a cardinalidade do maior conjunto de pontos que o algoritmo pode quebrar, o que significa que o algoritmo sempre pode aprender um classificador perfeito para qualquer rotulagem de pelo menos uma configuração desses pontos de dados. Foi originalmente definida por Vladimir Vapnik e Alexey Chervonenkis.

Figura 01 – Capacidade de Representação de uma Rede RBF.

$$\begin{array}{c}
 \Phi \\
 \left[\begin{array}{ccc} \exp\left(-\frac{\|\vec{x}_1 - \vec{c}_1\|^2}{2\sigma^2}\right) & \exp\left(-\frac{\|\vec{x}_1 - \vec{c}_2\|^2}{2\sigma^2}\right) & \dots & \exp\left(-\frac{\|\vec{x}_1 - \vec{c}_m\|^2}{2\sigma^2}\right) \\ \exp\left(-\frac{\|\vec{x}_2 - \vec{c}_1\|^2}{2\sigma^2}\right) & \exp\left(-\frac{\|\vec{x}_2 - \vec{c}_2\|^2}{2\sigma^2}\right) & \dots & \exp\left(-\frac{\|\vec{x}_2 - \vec{c}_m\|^2}{2\sigma^2}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \exp\left(-\frac{\|\vec{x}_p - \vec{c}_1\|^2}{2\sigma^2}\right) & \exp\left(-\frac{\|\vec{x}_p - \vec{c}_2\|^2}{2\sigma^2}\right) & \dots & \exp\left(-\frac{\|\vec{x}_p - \vec{c}_m\|^2}{2\sigma^2}\right) \end{array} \right] \begin{array}{c} \vec{w} \\ \left\{ \begin{array}{c} w_1 \\ w_2 \\ \vdots \\ w_m \end{array} \right\} \end{array} = \begin{array}{c} \vec{d} \\ \left\{ \begin{array}{c} d_1 \\ d_2 \\ \vdots \\ d_p \end{array} \right\} \end{array}
 \end{array}$$

$p \times m$ $m \times 1$

onde,

$p \dots$ número total de padrões de treinamento

$m \dots$ número total de neurônios na camada interna

$f(\vec{x}_i) = d_i$ para $i=1, 2, \dots, p$ $\vec{x}_i \in \mathbb{R}^m$ e $d_i \in \mathbb{R}$

Rede de Regularização:

$$\sigma = \frac{\text{dist}_{\max}}{\sqrt{2m}}$$

1. $m = p$

2. $\vec{w} = \Phi^{-1} \cdot \vec{d}$

3. Os centros \vec{c}_i são os próprios \vec{x}_i , ou seja, $\vec{c}_i = \vec{x}_i$ para $i=1, \dots, p$

Redes RBF

1. $m \ll p$

2. $\vec{w} = \Phi^+ \cdot \vec{d}$ para $\Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T$ (pseudo-inversa)

3. Os centros \vec{c}_i não coincide com os \vec{x}_i (há vários métodos para calculá-los)

Figura 02 – Cálculo dos Pesos de uma Rede de Regularização e de uma Rede RBF.

5.5 TEORIA DA REGULARIZAÇÃO

Determinação dos Coeficientes da Expansão

Introduzimos agora as seguintes definições:

$$F_\lambda = [F_\lambda(x_1), F_\lambda(x_2), \dots, F_\lambda(x_n)]^T \quad (21)$$

$$d = [d_1, d_2, \dots, d_n]^T \quad (22)$$

$$G = \begin{bmatrix} G(x_1, x_1) & G(x_1, x_2) & \dots & G(x_1, x_N) \\ G(x_2, x_1) & G(x_2, x_2) & \dots & G(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ G(x_N, x_1) & G(x_N, x_2) & \dots & G(x_N, x_N) \end{bmatrix} \quad (23)$$

$$W = [w_1, w_2, \dots, w_n]^T \quad (24)$$

As classes de funções de Green cobertas pelo teorema de Micchelli incluem *multiquádricas inversas* e *funções gaussianas*, mas não *multiquádricas*. Na prática, podemos sempre escolher λ suficientemente grande para garantir que $G + \lambda I$ seja definida positivamente e assim inversiva. Isto, por sua vez, significa que o sistema linear de Equações (27) terá uma única solução dada por

$$w = (G + \lambda I)^{-1} d \quad (28)$$

Assim, tendo selecionado o operador diferencial D e portanto tendo identificado a função de Green associada $G(x_j, x_i)$, onde $i = 1, 2, \dots, N$, podemos usar a Eq. (28) para obter o vetor de peso w para um vetor resposta desejada específico d e um valor apropriado de parâmetro de regularização λ . Concluindo, podemos afirmar que a solução do problema de regularização é dada pela expansão:

$$F_\lambda(x) = \sum_{i=1}^N \omega_i G(x, x_i) \quad (29)$$

Figura 03 – Cálculo dos Pesos de uma Rede de Regularização Considerando o Termo de Regularização de Tikonov.

Com base no exposto anteriormente, determine:

- Os pesos desta rede utilizando a Rede de Regularização de acordo com a Figura 02;
- Os pesos da Rede de Regularização de acordo com a Figura 03 e equação (28). Realizar 5 estudos de caso, a saber, para $\lambda=0$, $\lambda=1$, $\lambda=10$, $\lambda=100$ e $\lambda=1000$.
- Os pesos desta rede utilizando a Rede RBF com 14 neurônios na camada interna;
- Repita o item anterior com 25 neurônios na camada interna;
- Repita o item anterior com 50 neurônios na camada interna;
- Repita o item anterior com 80 neurônios na camada interna.

Note: gere os centros das redes RBF acima através do método aleatório.

Tabela 1 – Questionário Feito aos Senadores dos Estados Unidos em 1984.

| | |
|----|---|
| 1 | n,y,y,n,y,n,n,n,y,n,y,n,y, democrat |
| 2 | n,y,y,n,y,n,n,n,n,n,y,y,y, democrat |
| 3 | n,y,n,y,y,n,n,n,n,y,y,n,y, republican |
| 4 | y,y,y,n,n,y,y,n,y,n,n,n,y,y, democrat |
| 5 | y,y,y,n,n,y,y,n,n,n,n,n,y,y, democrat |
| 6 | y,n,y,n,n,y,y,y,n,n,n,n,y,y, democrat |
| 7 | y,n,y,n,n,y,y,n,y,n,n,n,y,y, democrat |
| 8 | y,y,y,n,n,y,y,n,y,n,n,n,y,y, democrat |
| 9 | y,n,n,y,n,y,y,n,n,y,y,n,y, republican |
| 10 | y,y,y,n,n,y,y,n,y,n,n,n,y,y, democrat |
| 11 | n,y,n,y,y,n,n,n,n,y,y,n,n, republican |
| 12 | y,y,y,n,n,y,y,y,n,n,y,n,y,y, democrat |
| 13 | n,y,n,y,y,n,n,n,n,y,y,n,y, republican |
| 14 | y,y,y,n,n,y,y,n,n,n,n,n,y,y, democrat |
| 15 | n,y,n,y,y,n,n,n,n,y,y,n,n, republican |
| 16 | y,y,n,y,y,n,n,n,n,n,y,n,y, republican |
| 17 | n,y,n,y,y,n,n,y,n,y,y,n,n, republican |
| 18 | y,n,y,n,n,y,y,y,y,n,y,n,y,y, democrat |
| 19 | y,n,y,n,n,y,y,n,n,n,n,n,y, democrat |
| 20 | y,n,y,n,n,y,y,n,n,n,n,n,y,y, democrat |
| 21 | y,y,y,n,n,y,y,n,n,n,n,n,n,y, democrat |
| 22 | y,y,y,n,n,y,n,n,n,n,n,y,n,y, democrat |
| 23 | y,y,y,n,n,y,y,n,y,n,n,n,y,y, democrat |
| 24 | n,y,n,y,y,n,n,n,n,y,y,n,y, republican |
| 25 | y,y,n,y,y,n,n,y,n,y,y,n,n, republican |
| 26 | n,y,n,y,y,n,n,y,y,y,y,n,n, republican |
| 27 | n,y,n,y,y,n,n,y,y,y,y,n,y, republican |
| 28 | n,y,n,y,y,n,n,y,n,y,y,n,y, republican |
| 29 | n,y,n,y,y,n,n,y,n,y,y,n,y, republican |
| 30 | n,y,n,y,y,n,n,n,n,y,y,n,n, republican |
| 31 | y,y,y,n,n,y,y,n,y,n,n,n,n,y, democrat |
| 32 | y,y,n,y,y,y,n,n,n,y,y,n,y, republican |
| 33 | n,y,n,y,y,y,n,n,y,y,y,n,y, republican |
| 34 | n,y,n,y,y,n,n,y,n,y,y,n,n, republican |
| 35 | y,y,y,n,n,y,y,n,n,n,n,n,y,y, democrat |
| 36 | y,y,y,y,n,y,y,y,y,n,n,y,n,y, republican |
| 37 | y,n,y,y,n,y,y,n,n,y,n,y, republican |
| 38 | y,n,y,n,y,y,y,y,y,n,n,y,y,y, democrat |
| 39 | n,y,y,y,y,n,n,y,y,n,y,n,n, democrat |
| 40 | n,y,y,y,y,n,y,y,y,y,y,n,y, democrat |
| 41 | y,y,y,n,y,n,n,y,y,n,y,n,y, democrat |
| 42 | n,n,n,y,n,n,n,y,n,y,y,n,n, republican |
| 43 | n,n,n,y,y,n,n,y,n,y,y,n,y, republican |
| 44 | n,n,n,y,y,n,n,n,n,y,y,n,n, republican |
| 45 | n,n,y,n,y,n,n,y,y,y,y,n,y, democrat |
| 46 | n,n,n,y,y,n,n,y,n,y,y,n,n, republican |
| 47 | n,n,n,y,y,n,n,n,n,y,y,n,n, republican |
| 48 | n,y,y,n,y,y,y,y,y,n,y,n,y, democrat |

| | |
|-----|---|
| 49 | y,n,y,n,n,y,y,y,n,n,n,n,y,y, democrat |
| 50 | y,n,y,n,n,y,y,y,n,n,n,y,y, democrat |
| 51 | y,n,y,n,n,y,y,y,n,n,n,y,y, democrat |
| 52 | y,n,y,n,y,n,n,n,n,n,n,n,y, democrat |
| 53 | y,n,n,n,y,y,n,y,y,n,n,y,n,y, democrat |
| 54 | y,y,y,n,y,y,y,y,n,n,n,n,y, democrat |
| 55 | y,n,n,n,y,n,n,n,y,y,n,y,n,y, democrat |
| 56 | y,n,y,n,y,y,n,n,y,n,n,y,n,y, democrat |
| 57 | y,y,y,n,n,n,y,n,y,n,n,n,y,y, democrat |
| 58 | n,y,n,y,y,n,n,n,n,y,y,y,n,y, republican |
| 59 | n,n,y,n,n,y,y,y,n,n,n,n,y,y, democrat |
| 60 | y,n,y,n,n,y,y,n,y,n,n,n,y,y, democrat |
| 61 | y,y,y,y,n,y,n,n,n,y,y,y,n,y, republican |
| 62 | n,y,y,n,n,n,y,y,y,n,n,n,y,y, democrat |
| 63 | n,n,n,y,y,n,n,n,n,y,y,y,n,n, republican |
| 64 | n,n,n,y,y,n,n,y,n,y,n,y,n,y, republican |
| 65 | n,n,y,n,y,n,y,y,n,n,n,y,n,y, democrat |
| 66 | n,n,n,y,y,n,n,y,n,y,y,y,n,y, republican |
| 67 | n,n,n,y,y,n,n,y,n,y,y,y,n,n, republican |
| 68 | n,y,n,y,y,n,n,y,y,y,y,n,n,y, republican |
| 69 | n,n,y,n,y,y,y,y,n,n,n,y,n,y, democrat |
| 70 | y,n,y,n,y,y,y,n,n,n,n,n,y,y, democrat |
| 71 | n,n,n,y,n,y,y,y,n,n,y,y,n,y, republican |
| 72 | n,n,n,y,y,y,y,y,n,y,y,y,n,y, republican |
| 73 | n,y,n,y,y,n,n,n,n,y,y,y,n,y, republican |
| 74 | n,n,n,n,n,y,y,y,n,y,y,y,y,y, democrat |
| 75 | n,y,n,y,y,n,n,y,y,y,y,y,n,y, republican |
| 76 | n,n,y,n,n,y,y,y,n,n,y,n,y,y, democrat |
| 77 | y,y,n,y,y,n,n,y,n,y,y,y,n,y, republican |
| 78 | n,y,y,n,y,n,y,y,y,n,y,n,y,y, democrat |
| 79 | n,n,y,n,y,y,y,y,y,n,y,y,n,y, democrat |
| 80 | n,y,n,y,y,n,n,n,n,y,y,y,n,n, republican |
| 81 | y,y,n,y,y,n,n,n,y,y,y,y,n,n, republican |
| 82 | y,y,n,y,y,y,n,n,n,y,y,y,n,n, republican |
| 83 | n,y,n,y,y,n,n,y,n,y,y,y,n,n, republican |
| 84 | n,y,n,n,y,n,n,n,n,y,y,y,y,y, democrat |
| 85 | n,n,n,n,y,y,n,n,n,y,y,y,n,y, democrat |
| 86 | n,y,y,n,y,y,n,n,y,y,y,y,n,y, democrat |
| 87 | n,y,n,y,y,y,n,n,n,y,y,y,n,y, republican |
| 88 | y,n,y,y,y,y,n,y,n,y,n,y,y,y, republican |
| 89 | y,n,y,y,y,y,n,y,y,y,n,y,y,y, republican |
| 90 | y,n,y,n,n,y,y,y,y,n,n,y,n,y, democrat |
| 91 | n,n,n,n,y,n,n,y,y,y,y,y,n,y, democrat |
| 92 | n,y,y,n,n,y,y,y,n,n,n,n,y,y, democrat |
| 93 | y,n,y,n,n,y,y,y,n,n,n,n,y,y, democrat |
| 94 | n,n,y,n,n,y,y,y,y,n,n,n,y,y, democrat |
| 95 | n,n,y,n,n,y,y,y,y,n,n,n,y,y, democrat |
| 96 | n,y,y,n,n,y,y,y,y,n,n,n,y,y, democrat |
| 97 | y,n,y,n,n,y,y,n,n,n,n,n,y,y, democrat |
| 98 | n,n,n,n,y,y,y,n,y,n,n,y,y,y, democrat |
| 99 | n,n,y,n,n,y,y,n,n,n,n,n,y,y, democrat |
| 100 | n,n,y,n,y,y,y,y,y,n,n,n,y,y, democrat |

Nota: na base de dados acima, o valor alfa-numérico “n” deve ser substituído pelo valor “0” e o valor “y” pelo número “1”. Observe que as Redes RBF só aceitam números nas suas entradas e saídas. O valor “republican” deve ser substituído por “1” e o valor “democrat” por “0”.

Descrição do Corpo do Relatório

1. **Capa** (Nome da Instituição, Nome da Disciplina, Título, Nome do Professor, Integrantes do Grupo e Data/local).
2. **Corpo do Relatório**
 - Introdução (máximo de uma página contendo enunciado do problema e outras informações relevantes que se acharem necessárias)
 - Objetivo (máximo de uma página)
 - Descrição técnica da metodologia utilizada para resolver o problema proposto
 - Análise e apresentação dos resultados (de acordo com o explicado na proposta dos parágrafos anteriores)
 - Conclusão
3. **Apêndice A**: um pequeno resumo e/ou descrição da linguagem utilizada
4. **Apêndice B**: listagem completa do código fonte

Boa Sorte ☺!

Prof. Paulo Marcelo Tasinaffo.
DCTA-ITA-IEC Divisão de Ciência da Computação.
Sala 107, TEL: +55 12 3947-6945.
e-mail: tasinaffo@ita.br.