

Trabalho 7 - Parte II

Nome: Gabriel Alessi Posonski

RA: 2259583

1 Introdução

Este é um relatório contendo as respostas solicitadas no Trabalho 7 - Parte II. O trabalho tem como objetivo usar o treinamento de uma rede *Multilayer Perceptron* (MLP) para estimar os pesos nas notas de provas e trabalhos de turmas de Algoritmos na Universidade Tecnológica Federal do Paraná (UTFPR). A base de dados utilizada foi fornecida na disciplina de Inteligência Artificial e contém notas de diversos alunos ao longo de vários semestres.

2 Metodologia

Essa seção descreve os parâmetros e dados utilizados para a obtenção dos resultados.

2.1 Preparação dos Dados

- Dataset Original: 196 amostras com variáveis PROVA1, PROVA2, TRABALHO e SITUAÇÃO.
- Normalização: MinMaxScaler aplicado às variáveis de entrada (intervalo [0,1]).
- Balanceamento: RandomUnderSampler reduzindo para 134 amostras (67 aprovados, 67 reprovados).
- Divisão: Holdout estratificado 2/3 treino (89 amostras) e 1/3 teste (45 amostras)

| Configurações Testadas | | |
|------------------------|-------------|--------------------------------------|
| Configuração | Arquitetura | Descrição |
| 1 | (5,) | 1 camada oculta com 5 neurônios |
| 2 | (10,) | 1 camada oculta com 10 neurônios |
| 3 | (5, 3) | 2 camadas ocultas (5 e 3 neurônios) |
| 4 | (10, 5) | 2 camadas ocultas (10 e 5 neurônios) |

2.2 Parâmetros de Treinamento

- Épocas testadas: 30, 50, 100
- Solver: Adam
- Learning rate: 0.01
- Total de experimentos: 12 (4 configurações \times 3 épocas)

3 Resultados

| Tabela com Resultados por configuração | | | | | | |
|--|--------------|-------------|--------|---------------------|--------------------|-----------|
| | Configuração | Arquitetura | Épocas | Acurácia Treino (%) | Acurácia Teste (%) | Convergiu |
| 1 | 1 | (5,) | 30 | 59.55 | 51.11 | Não |
| 2 | 1 | (5,) | 50 | 76.40 | 73.33 | Não |
| 3 | 1 | (5,) | 100 | 85.39 | 80.00 | Não |
| 4 | 2 | (10,) | 30 | 78.65 | 73.33 | Não |
| 5 | 2 | (10,) | 50 | 91.01 | 91.11 | Não |
| 6 | 2 | (10,) | 100 | 97.75 | 95.56 | Não |
| 7 | 3 | (5,3) | 30 | 50.56 | 48.89 | Não |
| 8 | 3 | (5,3) | 50 | 50.56 | 48.89 | Sim |
| 9 | 3 | (5,3) | 100 | 50.56 | 48.89 | Sim |
| 10 | 4 | (10,5) | 30 | 86.52 | 86.67 | Não |
| 11 | 4 | (10,5) | 50 | 92.13 | 91.11 | Não |
| 12 | 4 | (10,5) | 100 | 94.38 | 91.11 | Não |

| Melhores acurácias de Teste | | | | |
|-----------------------------|--------------|-------------|--------|--------------|
| | Configuração | Arquitetura | Épocas | Acurácia (%) |
| 6 | 2 | (10,) | 100 | 95.56 |
| 5 | 2 | (10,) | 50 | 91.11 |
| 11 | 4 | (10,5) | 50 | 91.11 |
| 12 | 4 | (10,5) | 100 | 91.11 |
| 10 | 4 | (10,5) | 30 | 86.67 |

3.1 Análise dos Resultados

Analisando os principais resultados, tem-se:

Configuração 2 (10,) - Vencedora

- **Melhor resultado geral:** 95,56% de acurácia no conjunto de teste após 100 épocas.
- **Progressão consistente:** A acurácia aumentou de forma constante com o número de épocas: 73,33% (30 épocas) → 91,11% (50 épocas) → 95,56% (100 épocas).
- **Capacidade de aprendizado:** Demonstrou forte capacidade de aprendizado ao longo do tempo, superando todas as outras configurações.
- **Overfitting moderado:** A diferença entre a acurácia de treino e teste foi de apenas 2,19%, considerado um valor aceitável.

Configuração 4 (10,5) - Segundo Lugar

- **Resultado:** Atingiu 91,11% de acurácia no teste tanto com 50 quanto com 100 épocas.
- **Estabilização precoce:** A performance não melhorou com o aumento de épocas, indicando um possível platô no aprendizado.

- **Complexidade excessiva:** Apesar de apresentar mais camadas, a arquitetura não superou o desempenho da Configuração 2.
- **Overfitting mais acentuado:** O aumento da complexidade resultou em uma diferença maior entre treino e teste.

Configuração 1 (5,) - Terceiro Lugar

- **Resultado:** Obteve 80,00% de acurácia no teste após 100 épocas.
- **Limitação de capacidade:** O número reduzido de neurônios provavelmente limitou a modelagem de padrões mais complexos.
- **Progressão linear:** Houve melhoria constante com o aumento de épocas, embora insuficiente frente às demais configurações.
- **Underfitting:** Indicativo de que a rede não conseguiu aprender suficientemente bem os padrões dos dados.

Configuração 3 (5,3) - Resultado Insatisfatório

- **Resultado:** Obteve apenas 48,89% de acurácia em todas as épocas testadas.
- **Convergência prematura:** A rede parou de aprender após 50 épocas, sem progresso posterior.
- **Performance abaixo do acaso:** A acurácia foi inferior à de uma escolha aleatória (50%), sugerindo problemas sérios na arquitetura.
- **Arquitetura inadequada:** A configuração provavelmente é demasiadamente restrita para a tarefa em questão.

Análise do Impacto das Épocas

Tabela 1: Acurácia no conjunto de teste em função do número de épocas

| Épocas | Config. 1 | Config. 2 | Config. 3 | Config. 4 |
|--------|-----------|---------------|-----------|-----------|
| 30 | 51,11% | 73,33% | 48,89% | 86,67% |
| 50 | 73,33% | 91,11% | 48,89% | 91,11% |
| 100 | 80,00% | 95,56% | 48,89% | 91,11% |

3.1.1 Importância das variáveis

A importância das variáveis foi determinada analisando os pesos da matriz de conexão entre a camada de entrada e a primeira camada oculta do modelo MLP vencedor. Para cada variável de entrada (Prova 1, Prova 2 e Trabalho), calcula-se a média dos valores absolutos de todos os pesos que conectam essa variável aos 10 neurônios da camada oculta. Os valores resultantes foram então normalizados proporcionalmente para somar 10 pontos, facilitando a interpretação em termos percentuais. Concluiu-se através do treinamento vencedor que os pesos de cada método avaliativo são:

- Prova 1: Peso 3,12 (31,2% de importância).

- Prova 2: Peso 3,98 (39,8% de importância).
- Prova 3: Peso 2,9 (29% de importância).

4 Conclusão

A análise comparativa das quatro configurações demonstra que o número de neurônios na camada oculta impacta fortemente a capacidade de aprendizado da MLP. A Configuração 2 (10,) mostrou-se a mais eficiente, equilibrando capacidade de generalização com baixo overfitting, e apresentando um crescimento consistente da acurácia com o número de épocas. A Configuração 4, apesar de mais complexa, não apresentou ganho em performance, sugerindo que redes mais profundas nem sempre são vantajosas para este problema. A Configuração 1 evidenciou limitação por simplicidade, enquanto a Configuração 3 falhou completamente em aprender os padrões esperados, reforçando a importância de uma boa escolha arquitetural.

Com o valor das importâncias das variáveis, pode-se concluir que o valor mais próximo do real peso das atividades avaliativas são: 3,12 para a prova 1, 3,98 para a prova 2 e 2,9 para o trabalho, totalizando 10. Com esse trabalho, foi possível compreender melhor o funcionamento da MLP e sua capacidade de geração de modelos preditivos. A utilização de dados reais trouxe também reflexões sobre como se deve tratar a informação antes de inseri-la em um treinamento de aprendizagem de máquina.