

Treinamento Supervisionado de Redes Neurais Artificiais para Predição da Qualidade de Vinhos Utilizando Backpropagation

Iuri Pedroso, Yan Gabriel, Herich Gabriel de Campos¹

Disciplina de Inteligência Artificial II

Curso de Ciência da Computação

Universidade Estadual do Centro-Oeste (Unicentro)

Guarapuava – PR – Brasil

Abstract

This project develops supervised neural network training using physicochemical data from red and white wine samples to predict their quality. The presented model uses the backpropagation algorithm for training, with the mean squared error as the cost function. Its pipeline includes data preprocessing, network architecture definition, supervised training, and performance evaluation. The overall results demonstrate the model's ability to perform its initial predictions, both in learning relevant patterns for estimating wine quality and highlighting the potential of neural networks for regression tasks in the context of food sensory analysis.

Resumo

Este trabalho é um projeto que desenvolve um treinamento supervisionado de redes neurais, onde se utiliza dados físico-químicos de amostras de vinho tinto e branco para predição de suas qualidades. O modelo apresentado utiliza o algoritmo de backpropagation para seu treinamento, tendo como função de custo o erro quadrático médio. Sua pipeline inclui etapas de pré-processamento dos dados, definição da arquitetura da rede, treinamento supervisionado e avaliação do desempenho. A representação geral em seus resultados, demonstram a capacidade do modelo em sua predição inicialmente dita, tanto em sua parte de aprender padrões relevantes para estimar a qualidade dos vinhos, quanto evidenciar o potencial das redes neurais em tarefas de regressão no contexto de análise sensorial de alimentos.

¹Graduandos em Ciência da Computação – Universidade Estadual do Centro-Oeste
iuripedroso11072005@gmail.com, herichop013l@gmail.com, yangabrielreis@gmail.com

1. Introdução

“A avaliação da qualidade do vinho é uma tarefa complexa, que depende de múltiplos fatores físico-químicos e sensoriais, sendo tradicionalmente realizada por especialistas treinados.” (Cortez et al., 2009). Com o avanço das técnicas de inteligência artificial, métodos baseados em redes neurais artificiais têm se mostrado promissores para automatizar e tornar mais precisa a predição da qualidade de produtos alimentícios, uma vez que, a partir dos dados, é possível analisar requisitos e características que fazem um produto se sobressair em relação aos demais. Entre esses métodos, temos o Perceptron Multicamadas (MLP), que se destaca pela sua capacidade de lidar com relações não lineares, extraíndo o padrão em sua taxa de aprendizagem. Neste trabalho, é proposto a implementação e avaliação de uma MLP para prever a qualidade dos vinhos brancos e tintos fornecidos pelo dataset utilizado.

2. Desenvolvimento

Neste trabalho, utilizamos o conjunto de dados disponibilizado por <https://archive.ics.uci.edu/dataset/186/wine+quality>, dessa forma, o dataset contém informações físico-químicas de amostras de vinho tinto e branco, como acidez, teor alcoólico, pH, concentração de açúcar, entre outros, além da nota de qualidade atribuída por especialistas em uma escala de 0 a 10. Os dados disponibilizados permitem uma análise séria do conteúdo para predição e treinamento da rede.

A rede neural foi proposta a partir do tipo Perceptron Multicamadas (MLP, do inglês Multilayer Perceptron) (Rumelhart, Hinton e Williams, 1986). A MLP é uma arquitetura de rede neural artificial composta por múltiplas camadas de neurônios, capaz de aprender relações não lineares complexas entre as variáveis de entrada e saída. Nesse contexto a MLP analisa os dados do dataset, tal qual, teor alcoólico, pH e concentração de açúcar, e busca prever a nota dos vinhos com a qualidade correspondentemente dita no dataset, assim, o processo de aprendizado ocorre por meio da propagação direta (forward propagation), onde os dados percorrem as camadas da rede até gerar uma previsão (predição), e da retropropagação do erro (backpropagation) que ajusta os pesos da rede para minimizar a diferença entre a previsão e o valor real.

2.1 Tratamento de dados

Para evitar com que o algoritmo acabe se confundindo com os atributos do vinho que são definidos em escalas diferentes, utilizamos a função StandardScaler para realizar a padronização dos dados. Essa função serve para definir que cada atributo tenha média igual a zero e desvio padrão igual a 1. Essa etapa é essencial para que todos os atributos seguissem um padrão e nenhum simplesmente dominasse a etapa de aprendizado do algoritmo, evitando assim um bias nos resultados que pendesse a um atributo específico.

Após isso, o conjunto de dados foi dividido em dois subconjuntos, um para treinamento e outro para teste. Para esta definição seguimos o padrão da biblioteca utilizada na função `train_test_split`, onde não alteramos os valores, logo foi definido automaticamente a seguinte divisão: 80% para treinamento e 20% para teste.

2.2 Definição de camadas e neurônios

A nossa arquitetura para o modelo MLP foi definida da seguinte maneira: 11 neurônios na camada de entrada, um para cada atributo físico-químico do vinho; Em relação às camadas ocultas, definimos uma única camada com 32 neurônios; Para a camada de saída, ela possui um único neurônio, pois queremos como saída um único valor, a predição da nota de qualidade do vinho.

2.3 Parâmetros de treinamento

O treinamento da rede foi configurado para ser executado ao longo de 1.000 épocas. Um aspecto particular do nosso experimento foi a definição da taxa de aprendizado (learning rate), que não foi um valor fixo. Em vez disso, para cada execução do treinamento, um valor foi sorteado aleatoriamente no intervalo entre 0.001 e 0.1.

2.4 Plotagem dos resultados e Output

Após o processo de treinamento, é necessário avaliar o desempenho testando o output da última camada da MLP. Para realizar esta tarefa utilizamos o erro quadrático médio (Mean Squared Error, ou MSE), o que penaliza os erros maiores encontrados nas comparações de forma quadrática, ou seja, ele não mede diretamente a distância dos erros, mas sim elevado ao quadrado, eliminando valores negativos e aumentando ainda mais os erros grandes.

Tendo esses resultados em mão, consideramos o acerto da nossa máquina como uma métrica aproximada, neste caso, definimos uma margem de tolerância de ± 0.5 em relação a nota real atribuída ao vinho.

Outra métrica importante da qual utilizamos para saída é a perda acumulada a partir das épocas. Utilizamos essa informação para plotar um gráfico de curva de perda utilizando a biblioteca matplotlib, onde podemos encontrar a convergência de nosso modelo ao decorrer do treinamento.

2.5 Bibliotecas utilizadas

O projeto foi desenvolvido em linguagem Python, com o suporte das seguintes bibliotecas: NumPy para operações numéricas, Scikit-learn para as ferramentas de pré-processamento e avaliação, e Matplotlib para a geração dos gráficos de resultados. O código-fonte completo do projeto pode ser acessado em: Herich Gabriel de Campos. *IA-2: Treinamento de MLP para predição da qualidade de vinhos*. Disponível em: <https://github.com/herich0/IA-2>.

3. Resultados

Os experimentos foram conduzidos utilizando o conjunto completo de dados de vinhos brancos e tintos (Wine Quality Dataset, Cortez et al., 2009), sendo 70% reservado para treinamento e 30% para teste. A arquitetura adotada foi uma MLP com 11 neurônios na camada de entrada, uma camada oculta contendo 32 neurônios e uma camada de saída com um único neurônio. O treinamento foi realizado ao longo de 1000 épocas, com taxa de aprendizado sorteada aleatoriamente em cada execução entre 0.001 e 0.1.

A Tabela 1 apresenta os principais resultados obtidos para os dois subconjuntos do dataset:

tabela 1 - Desempenho da MLP na predição da qualidade do vinho

Execução	número de camadas ocultas	número de neurônios nas camadas ocultas	Taxa de Aprendizagem	precisão p/ vinho tinto (margem de 0.5)	precisão p/ vinho branco (margem de 0.5)
1	1	32	0.1	56.88%	53.16%
2	1	32	0.01	56.56%	50.92%
3	3	64/32/16	0.01	41.25%	44.08%
4	3	64/32/16	0.1	58.13%	53.67%
5	2	128/64	0.1	57.50%	56.84%
6	2	32/32	0.1	57.19%	55.51%
7	2	32/16	0.1	58.13%	55.41%
8	1	16	0.1	58.44%	52.86%
9	1	16	0.2	56.88%	52.96%
10	2	128/64	0.01	55.00%	51.53%

As Figuras 1 e 2 mostram a curva de perda (loss) ao longo das épocas de treinamento para os dois datasets. Nota-se que, em ambos os casos, a função de custo apresenta comportamento decrescente e tende à estabilização, demonstrando que a rede foi capaz de aprender os padrões presentes sem sinais evidentes de sobreajuste.

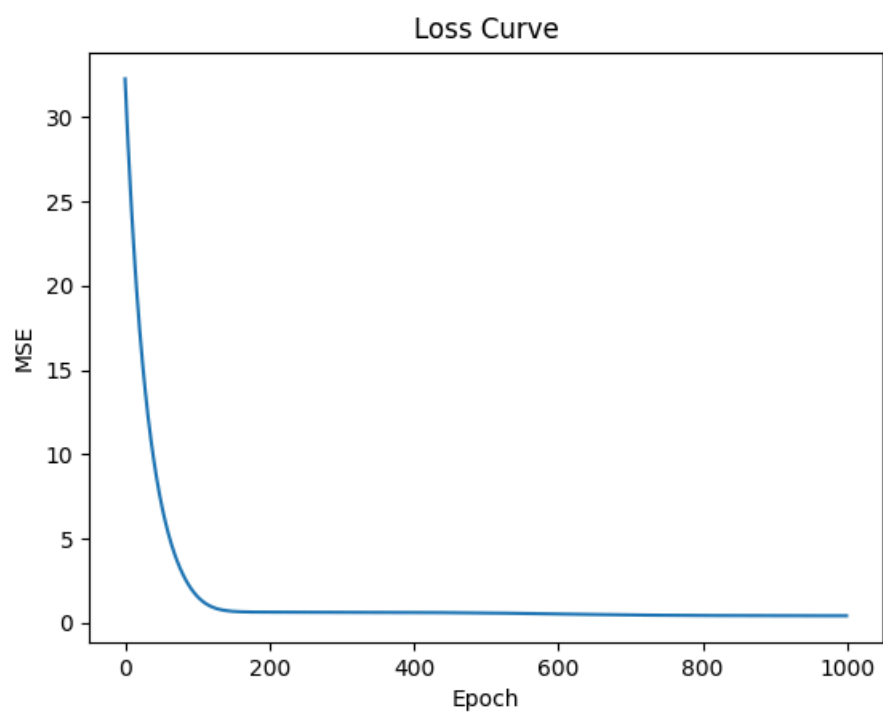


Figura 1: Curva de perda para o dataset de vinhos tintos.

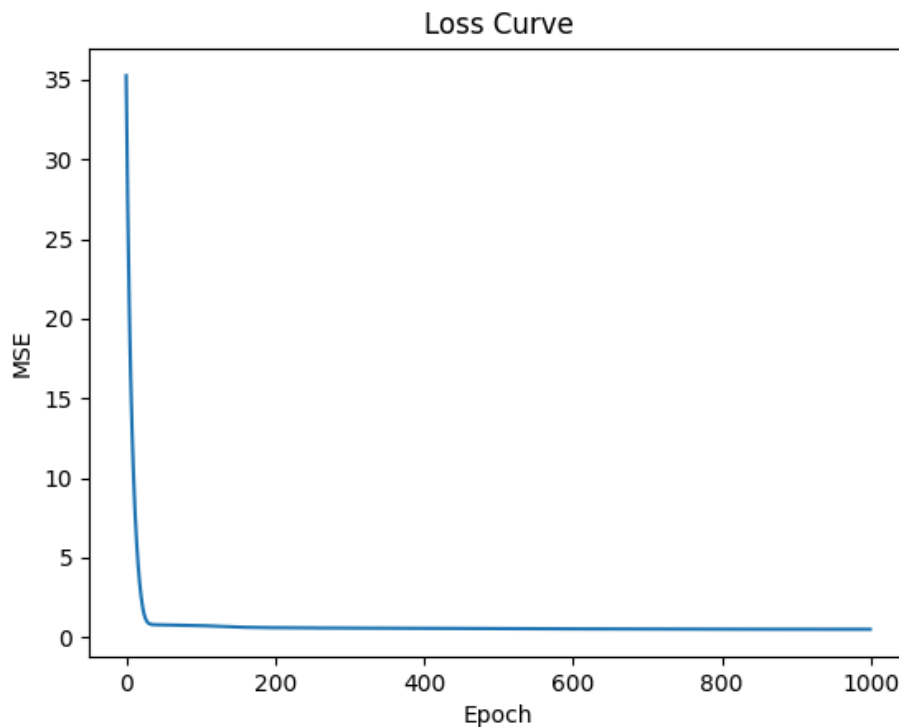


Figura 2: Curva de perda para o dataset de vinhos brancos.

Embora a precisão esteja em torno de 54% considerando a tolerância de ± 0.5 pontos, os resultados indicam que a MLP conseguiu capturar relações relevantes entre as variáveis físico-químicas e a qualidade dos vinhos, mostrando um desempenho razoável para uma arquitetura simples e com hiperparâmetros sorteados de forma aleatória.

4. Conclusão

Este trabalho apresentou o desenvolvimento e avaliação de uma rede neural do tipo Perceptron Multicamadas (MLP) para a predição da qualidade de vinhos com base em características físico-químicas. Os resultados obtidos mostram que, mesmo com uma arquitetura simples e hiperparâmetros definidos aleatoriamente, a rede conseguiu atingir um desempenho satisfatório, com MSE abaixo de 0.52 e precisão próxima de 54% em ambos os datasets.

A etapa de padronização dos dados demonstrou-se essencial para garantir que os diferentes atributos contribuíssem de forma equilibrada para o aprendizado. Além disso, a análise das curvas de perda confirma que o treinamento atingiu estabilidade, sem sinais evidentes de overfitting.

Como trabalhos futuros, sugere-se:

- Otimizar os hiperparâmetros utilizando técnicas como grid search ou random search;
- Experimentar arquiteturas mais profundas e otimizadores adaptativos (Adam, RMSprop);
- Avaliar o uso de técnicas de regularização, como dropout, para melhorar a generalização;
- Ampliar as métricas de avaliação, incluindo coeficiente de determinação (R^2) e validação cruzada.

Essas melhorias podem contribuir para aumentar a precisão do modelo e explorar melhor o potencial das redes neurais na análise sensorial de alimentos.

Referências

- Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). *Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems.*
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors.*
- Herich Gabriel de Campos, Iuri Pedroso, Yan Gabriel. *IA-2: Treinamento de MLP para predição da qualidade de vinhos.* Disponível em: <https://github.com/herich0/IA-2>