

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**REDES NEURAIS ARTIFICIAIS
ATIVIDADE 2 - PARTE 1: Overfitting e Underfitting**

Aluna: Priscila Aparecida Dias Nicácio

Professores: Prof. Antônio Braga e Prof. Frederico Gualberto.

Belo Horizonte, Agosto de 2025

Sumário

1.	INTRODUÇÃO.....	3
2.	FUNDAMENTAÇÃO TEÓRICA.....	3
3.	RESOLUÇÃO DA QUESTÃO	3
3.1	QUESTÃO 1A	4
3.2	QUESTÃO 1B	4
3.3	QUESTÃO 1C	5
3.4	QUESTÃO 1D.....	5
4	CONCLUSÃO.....	5
5	REFERÊNCIAS BIBLIOGRÁFICAS.....	6

1. INTRODUÇÃO

Em problemas de regressão, o objetivo principal é construir modelos que descrevam a relação entre variáveis de entrada e saída, permitindo não apenas explicar os dados observados, mas também realizar previsões sobre novos cenários. No entanto, a qualidade desse processo depende da escolha adequada da complexidade do modelo, que influencia diretamente sua capacidade de generalização (BISHOP, 2006; HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

2. FUNDAMENTAÇÃO TEÓRICA

A escolha da complexidade do modelo envolve o equilíbrio entre underfitting e overfitting. O underfitting ocorre quando o modelo é simples demais, incapaz de representar a estrutura real dos dados, resultando em alto erro tanto no conjunto de treinamento quanto em novos dados. Um exemplo típico é o uso de modelos lineares para tentar ajustar funções não lineares (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; MONTGOMERY; PECK; VINING, 2012). O overfitting, por outro lado, ocorre quando o modelo é excessivamente complexo, ajustando-se inclusive ao ruído presente nos dados de treinamento. Isso leva a um baixo erro de treinamento, mas piora o desempenho em dados novos, pois o modelo não aprendeu o padrão real. O ponto ótimo situa-se em modelos intermediários, que capturam a tendência geral dos dados sem se prender a detalhes irrelevantes, garantindo melhor capacidade de generalização (KOHAVI; JOHN, 1997).

3. RESOLUÇÃO DA QUESTÃO

A tarefa sobre overfitting e underfitting foca no entendimento de conceitos e a relação com a dimensão do modelo e erros de treinamento e teste (GOLUB; VAN LOAN, 2013).

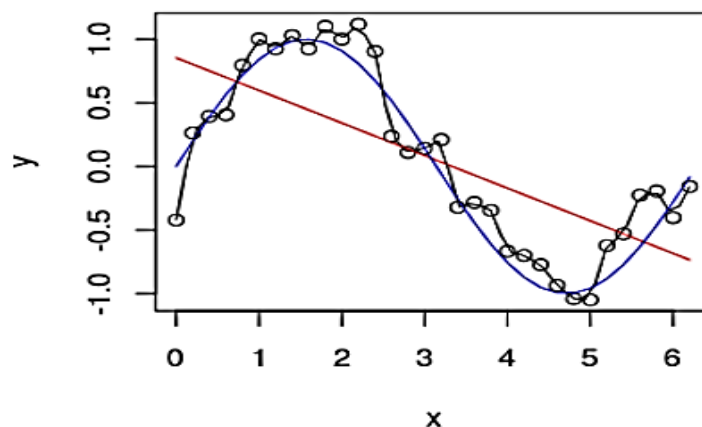


Figura 2: Ajuste de três modelos para um problema de regressão.

Desse modo, considerando-se a Figura 2, que apresenta os dados de treinamento para um problema de regressão, na imagem, tem-se que:

- Preto: pontos de dados (observações).
- Vermelho: modelo linear simples.
- Azul: modelo mais complexo, que “passa por quase todos os pontos”.

Na Tabela 1, abaixo há o resumo do que ocorre na figura 2.

Modelo	Treinamento	Teste/Novos Dados	Tipo de ajuste
Vermelho	Alto erro	Melhor que azul	Underfitting
Azul	Baixo erro	Pior que vermelho	Overfitting
Preto (pontos)	–	–	Dados reais

Tabela 1: Resumo do gráfico da figura 2 com o ajuste dos três modelos para o problema de regressão.

3.2 QUESTÃO 1A

Qual dos modelos parece melhor aproximação da função geradora, considerando ruído?

O modelo azul segue rigidamente cada ponto, é muito flexível, ajustando-se não só ao padrão real, mas também ao ruído. O modelo vermelho é muito simples, não captura a curvatura dos dados, caracterizando underfitting. Nenhum modelo é perfeito, mas o intermediário ideal - em que o azul exagera e o vermelho subestima - é aquele que segue a tendência geral dos dados sem se prender ao ruído. Logo, o modelo que captura a curva geral sem exagerar seria considerado a melhor aproximação da função geradora.

3.3 QUESTÃO 1B:

Qual apresenta menor erro de treinamento?

O modelo azul passa muito próximo de todos os pontos, então o erro de treinamento é mínimo (quase zero). O modelo vermelho possui erro maior, pois não consegue seguir a curva.

3.4 QUESTÃO 1C:

Qual deve ter melhor desempenho em dados novos?

- Vermelho: pode ser muito simples e perder padrões importantes (erro maior).
- Azul: ajusta demais ao ruído, tende a ter alto erro em novos dados (overfitting).

Um modelo intermediário teria melhor desempenho geral, mas entre os mostrados, o vermelho provavelmente generaliza melhor que o azul, embora ainda possa ser subótimo - conforme estudado na literatura (KOHAVI; JOHN, 1997).

3.5 QUESTÃO 1D:

Efeitos do dimensionamento do modelo e ajuste aos dados:

- Modelo simples (underfitting): não captura a complexidade da função (alto erro de treinamento e teste).
- Modelo complexo (overfitting): captura ruído (baixo erro de treinamento, mas alto erro em dados novos).

Erro de treinamento baixo não implica em bom desempenho a longo prazo: é ilusório se o modelo está se ajustando ao ruído, e não ao padrão real.

Regra prática: deve-se buscar modelo com complexidade equilibrada, pois captura a tendência dos dados sem memorizar ruído – conforme literatura (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; MONTGOMERY; PECK; VINING, 2012).

4. CONCLUSÃO

A atividade permitiu avaliar o impacto da complexidade do modelo no ajuste aos dados e na capacidade de generalização. A partir da Figura 2, observou-se que o modelo azul, altamente flexível, ajusta-se rigidamente a quase todos os pontos de dados, incluindo o ruído. Isso caracteriza overfitting, resultando em erro de treinamento mínimo, mas pior desempenho em dados novos. O modelo vermelho, linear simples, não consegue capturar a curvatura dos dados, caracterizando underfitting - ele apresenta maior erro de treinamento, mas tende a generalizar melhor que o modelo azul, embora ainda subestime a função geradora.

O modelo ideal seria intermediário, seguindo a tendência geral dos dados sem se prender aos detalhes aleatórios, equilibrando ajuste e generalização.

Os resultados reforçam conceitos fundamentais: erro de treinamento baixo não garante boa capacidade preditiva, e modelos excessivamente complexos podem memorizar o ruído, enquanto modelos muito simples podem ignorar padrões importantes. A análise evidencia que a escolha do modelo deve sempre buscar um compromisso entre ajuste e complexidade, garantindo que o modelo capture corretamente a função geradora e mantenha robustez frente a novos dados (BISHOP, 2006; HASTIE; TIBSHIRANI; FRIEDMAN, 2009; KOHAVI; JOHN, 1997).

5. REFERÊNCIAS BIBLIOGRÁFICAS

1. BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
2. HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009.
3. MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to Linear Regression Analysis*. 5. ed. Hoboken: Wiley, 2012.
4. KOHAVI, R.; JOHN, G. H. *Wrappers for Feature Subset Selection*. *Artificial Intelligence*, v. 97, n. 1-2, p. 273–324, 1997.
5. GOLUB, G. H.; VAN LOAN, C. F. *Matrix Computations*. 4. ed. Baltimore: Johns Hopkins University Press, 2013.