

Análise da Regressão Linear Múltipla para Eficiência Energética de Edifícios

Fernando Andrade Lima Tavares

3 de Junho de 2025

1 Introdução

Este relatório apresenta a análise do conjunto de dados de eficiência energética de edifícios e a aplicação de regressão linear múltipla para prever a carga térmica de aquecimento (*Heating Load*). A análise inclui a descrição do dataset, a formulação do problema de mínimos quadrados e a interpretação dos resultados gerados pelo modelo.

2 Descrição do Dataset

O dataset contém 768 amostras, cada uma representando um edifício simulado com diferentes características. Os atributos considerados são:

- X1: **Compacidade Relativa** - Mede a eficiência térmica do edifício.
- X2: **Área Superficial** - Influencia a troca de calor com o ambiente.
- X3: **Área das Paredes** - Relacionada à retenção de calor.
- X4: **Área do Telhado** - Impacta na absorção de calor.
- X5: **Altura Total** - Pode afetar a circulação de ar e a dispersão de calor.
- X6: **Orientação** - Determina a exposição ao sol.
- X7: **Área de Vidros** - Influencia o ganho/perda de calor.
- X8: **Distribuição da Área de Vidros** - Define a localização dos vidros no edifício.

Esses atributos constituem as variáveis independentes (vetor X), enquanto a variável dependente y corresponde à carga térmica de aquecimento (*Heating Load*), que o modelo busca prever.

O objetivo é prever a variável **Heating Load** (Carga Térmica de Aquecimento) com base nesses atributos.

3 Formulação do Problema de Mínimos Quadrados

A regressão linear múltipla busca encontrar um vetor de coeficientes β tal que minimize o erro quadrático médio entre os valores preditos \hat{y} e os valores reais y . A função de erro é definida como:

$$F(\beta) = \frac{1}{2m} \sum_{i=1}^m (y_i - X_i\beta)^2 \quad (1)$$

Onde:

- m é o número de amostras.
- y_i é o valor real da carga térmica.
- \hat{y}_i é a predição do modelo.

O vetor de predição é dado por:

$$\hat{y} = X\beta \quad (2)$$

E a solução otimizada para β pode ser obtida pela equação normal:

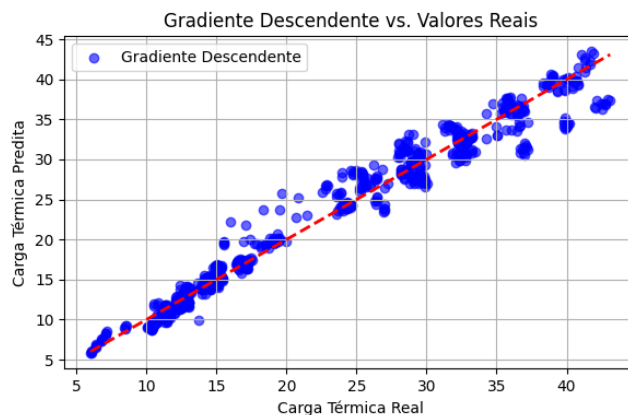
$$\beta = (X^T X)^{-1} X^T y \quad (3)$$

No código foi utilizado o *SGDRegressor*, que aproxima essa solução usando o método de Gradiente Descendente Estocástico (SGD).

4 Análise dos Resultados

A predição da carga térmica foi comparada aos valores reais através do gráfico de dispersão. A linha vermelha representa a relação ideal onde $\hat{y} = y$. Observa-se:

Figure 1: Gráfico de dispersão entre os valores reais da carga térmica e as predições do modelo (SGDRegressor)..



O modelo apresenta uma distribuição bem ajustada em torno da reta ideal, indicando uma boa precisão, algumas pequenas discrepâncias sugerem que características adicionais ou técnicas de regularização podem melhorar o desempenho do modelo. Como medida quantitativa de desempenho, foi utilizada a métrica de Erro Médio Quadrático (MSE), que avalia a proximidade entre os valores reais e preditos.

5 Conclusão

A regressão linear múltipla demonstrou desempenho satisfatório na tarefa de predição da carga térmica de aquecimento, capturando relações lineares relevantes entre os atributos.