Regressão Linear Múltipla Aplicada à Estimativa da Carga Térmica de Aquecimento

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 construtivas. O objetivo do conjunto de dados é modelar a demanda térmica para aquecimento, com base em variáveis físicas e arquitetônicas.

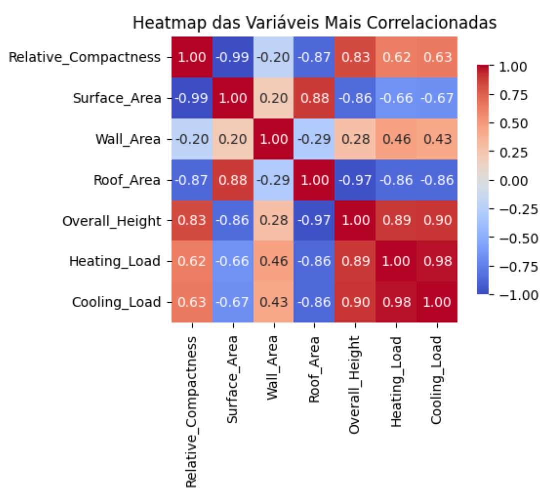
Os atributos considerados são:

- 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.
- X9: **Demanda de Resfriamento** – Representa a carga térmica de resfriamento do edifício.

Análise da Matriz de Correlação Compacta

Para focar nas associações mais relevantes com a carga de aquecimento, foram filtrados para que apenas as variáveis cuja correlação bivariada com *Heating_Load* apresenta magnitude $|r| \geq 0,40$. No heatmap compacto da Figura 1 observa-se que Overall_Height detém a correlação positiva mais forte (r+0,89), enquanto Area apresenta a correlação negativa mais intensa (r–0,86), indicando que coberturas mais amplas aliviam substancialmente a demanda de aquecimento. Em seguida, Relative_Compactness (r+0,62) e Surface_Area (r–0,66) exibem associações moderadas, refletindo o papel combinado da forma e da extensão de superfície na troca de calor. A Wall_Area aparece com correlação moderada positiva (r+0,46). Essas relações evidenciam as variáveis candidatas a maiores efeitos na previsão.

Figure 1: Heatmap das colunas com relação a *Heating Load*.



3 Formulação do Problema de Mínimos Quadrados

A regressão linear múltipla busca encontrar um vetor de coeficientes β tal que a função de custo (ou erro) quadrático médio entre as predições $\hat{y}$ e os valores reais y seja minimizada. A função objetivo é definida por:

$$F(\beta) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - x_i^\top \beta)^2 = \frac{1}{2m} \|y - X\beta\|^2 \tag{1}$$

Onde:

- m é o número de amostras.

- $x_i^\top$ é o vetor linha correspondente à i-ésima observação de atributos.
- X é a matriz de dados com dimensão $m \times n$.
- y é o vetor de saídas reais, com dimensão $m \times 1$.
- $\hat{y} = X\beta$ representa o vetor de predições do modelo.

Para encontrar os coeficientes β que minimizam a função de custo, derivamos $F(\beta)$ em relação a β:

$$\nabla_\beta F(\beta) = \frac{1}{2m} \nabla_\beta (y - X\beta)^\top (y - X\beta) \tag{2}$$

$$= \frac{1}{2m} [-2X^\top (y - X\beta)] \tag{3}$$

$$= -\frac{1}{m} X^\top (y - X\beta) \tag{4}$$

Igualando o gradiente a zero, obtemos a condição de mínimo:

$$X^\top X\beta = X^\top y \tag{5}$$

Assumindo que $X^\top X$ é inversível, a solução analítica (chamada de equação normal) é dada por:

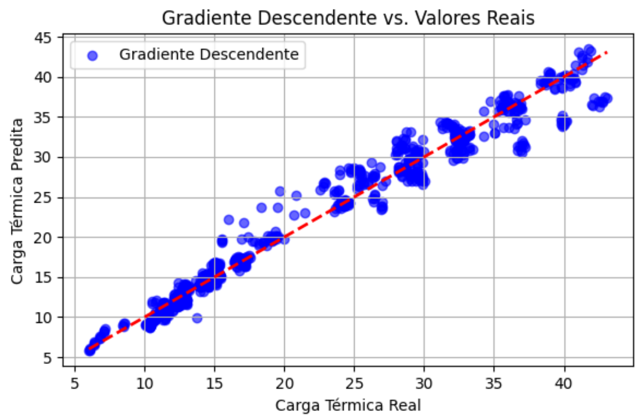
$$\beta = (X^\top X)^{-1} X^\top y \tag{6}$$

No código, optou-se por utilizar o *SGDRegressor* da biblioteca *scikit-learn*, que aproxima a solução por meio do método de Gradiente Descendente Estocástico, ajustando iterativamente os coeficientes β com base em amostras individuais.

4 Análise dos Resultados

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

Figure 2: Gráfico de dispersão entre os valores reais da carga térmica e as predições do modelo (SGDRegressor).



O modelo apresenta uma distribuição ajustada em torno da reta ideal, indicando que as predições acompanham satisfatoriamente os valores reais observados. A métrica quantitativa utilizada para avaliar o desempenho foi o Erro Médio Quadrático (MSE), que mede a média dos quadrados dos erros entre as predições e os valores reais, fornecendo uma indicação da precisão do modelo.

5 Conclusão

A regressão linear múltipla demonstrou desempenho satisfatório na tarefa de prever a demanda térmica de aquecimento dos edifícios, capturando relações lineares relevantes entre os atributos estruturais.

6 Referências

- **Scikit-learn:** Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 12, 2825–2830. Disponível em: `https://scikit-learn.org/`

- **Dataset – Energy Efficiency:** Tsanas, A., & Xifara, A. (2022). *Energy Efficiency Data Set.* Disponível no Kaggle: `https://www.kaggle.com/datasets/ujjwalchowdhury/energy-efficiency-data-set`

- **Fundamentos de Regressão Linear:** Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning.* Springer. Disponível em: `https://web.stanford.edu/~hastie/ElemStatLearn/`