



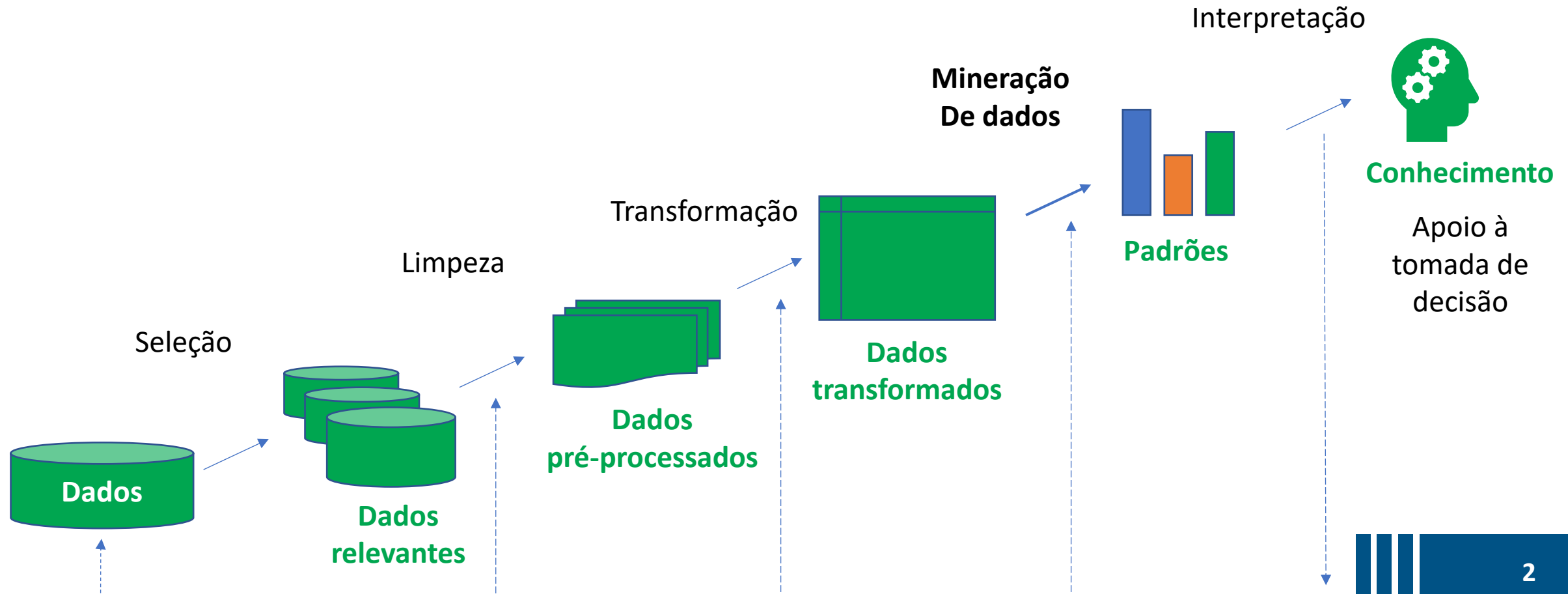
# Regressão linear

Ciência de Dados II

**Professor: Gabriel Machado Lunardi**  
gabriel.lunardi@ufsm.br

# O processo de KDD

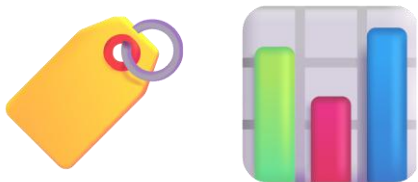
“É um processo de várias etapas, não trivial, **interativo** e **iterativo**, para a identificação de **padrões** válidos, novos e potencialmente úteis a partir de um grande conjunto de dados” (FAYYAD, 1996).



# Tarefas de aprendizado de máquina

---

- ✓ Aprendizado supervisionado
  - ✓ Predição
    - ✓ Classificação
    - ✓ Regressão
- ✓ Aprendizado não supervisionado
  - ✓ Agrupamento
  - ✓ Associação
- ✓ Aprendizado semi-supervisionado
- ✓ Aprendizado por reforço



## Supervisionado Classificação

Imagine **um professor supervisionando os alunos.**  
**O professor fornece exemplos e as respostas corretas.**  
O aluno aprende com esses exemplos para responder a **novas perguntas semelhantes.**

- ✓ Detectar se um e-mail é spam ou não.
  - ✓ spam/não spam.
- ✓ Identificar a raça de um cachorro em uma imagem.
  - ✓ Poodle, Labrador, etc.
- ✓ Diagnosticar um paciente como tendo gripe ou não.
  - ✓ gripe/saudável.



## Supervisionado Regressão

Prever um valor numérico contínuo.

- ✓ Prever o preço de um imóvel com base no tamanho, localização..
- ✓ Estimar o número de vendas de um produto no próximo mês.
- ✓ Prever a temperatura máxima para amanhã.

# Supervisionado

## Regressão

organização das técnicas



### Relação entre variáveis:

- ✓ Regressão linear simples e múltipla
- ✓ Regressão polinomial
- ✓ Regressão não-linear

### ✓ Quantidade de variáveis dependentes:

- ✓ Regressão univariada
- ✓ Regressão multivariada – prever a temperatura e a pressão atmosférica com base na altitude

### ✓ Técnicas

- ✓ Regressão Ridge
- ✓ Regressão Lasso
- ✓ Regressão Elastic Net
- ✓ Regressão por redes neurais
- ✓ Regressão Bayesiana
- ✓ Dentre várias outras



## Não Supervisionado



Imagine um grupo de crianças brincando livremente e se organizando sozinhas em grupos, sem um adulto ditando as regras.

- ✓ Segmentar clientes em diferentes grupos com base em seus hábitos de compra.
- ✓ Agrupar documentos por tópico.
- ✓ Redução de dimensionalidade para visualização de dados.



## Semi Supervisionado



Uma mistura dos dois anteriores!

Imagine um quebra-cabeça parcialmente resolvido como guia.

Usa uma **pequena quantidade de dados rotulados** e **uma grande quantidade de dados não rotulados** para treinar o modelo.

- ✓ Classificação de imagens da web, onde apenas algumas imagens têm **rótulos**.
- ✓ Análise de sentimentos em textos com poucos **exemplos rotulados**.





## Por reforço

Imagine treinar um cachorro com recompensas e punições.

O modelo aprende a tomar decisões interagindo com um ambiente.

- ✓ Robôs aprendendo a navegar em um labirinto.
- ✓ Sistemas de recomendação que aprendem as preferências do usuário.
- ✓ Algoritmos para jogar jogos como xadrez.

# Regressão linear

# Regressão linear

A regressão linear é uma técnica estatística usada para encontrar **a relação entre variáveis**. Em um contexto de ML, a regressão linear encontra a relação entre **atributos** e **um rótulo**.

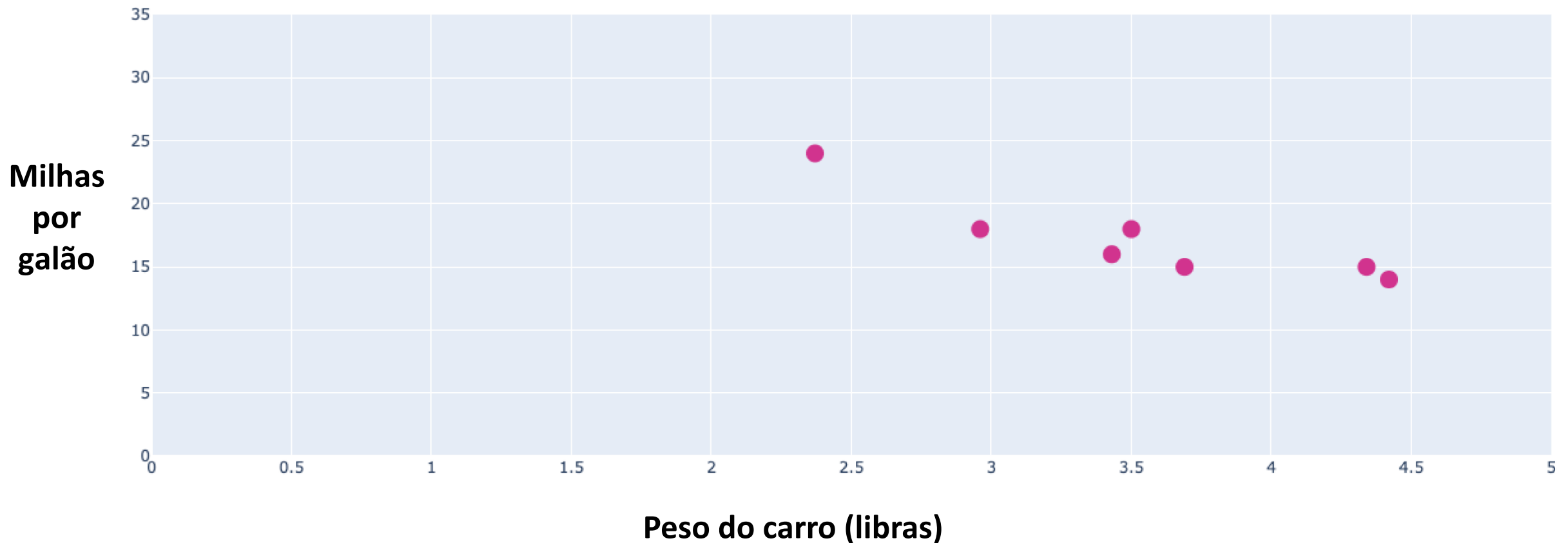
Peso do carro – libras em 1.000 (atributo)	Milhas por galão (rótulo)
3.5	18
3,69	15
3,44	18
3,43	16
4,34	15
4,42	14
2,37	24

prever a eficiência de combustível de um carro em milhas por galão com base no peso do carro



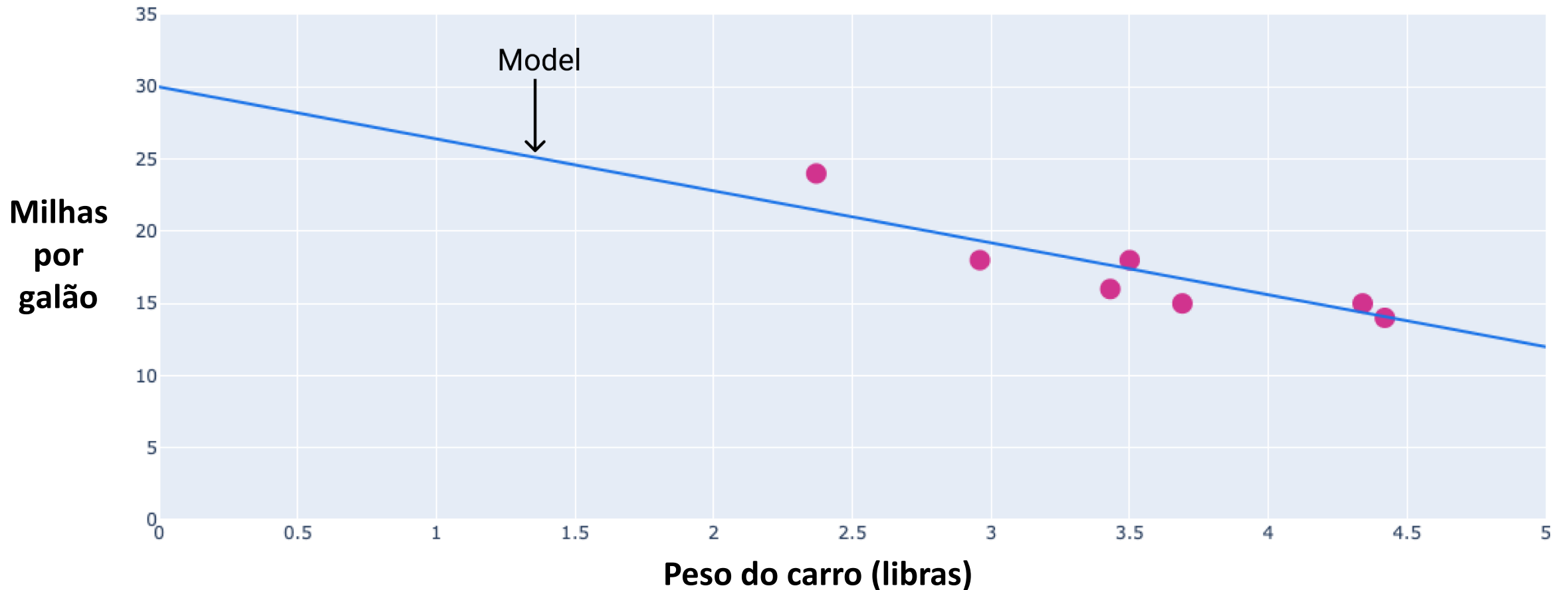
# Regressão linear

À medida que um carro fica mais pesado, a classificação de milhas por galão geralmente diminui. Logo, o carro fica mais “gastão”.



# Regressão linear

Podemos “desenhar” uma linha de ajuste a partir dos pontos (dados).  
Essa linha, damos o nome de modelo.



# Definição da Regressão linear

✓ Em termos algébricos, o modelo seria definido como:

$$y = mx + b$$

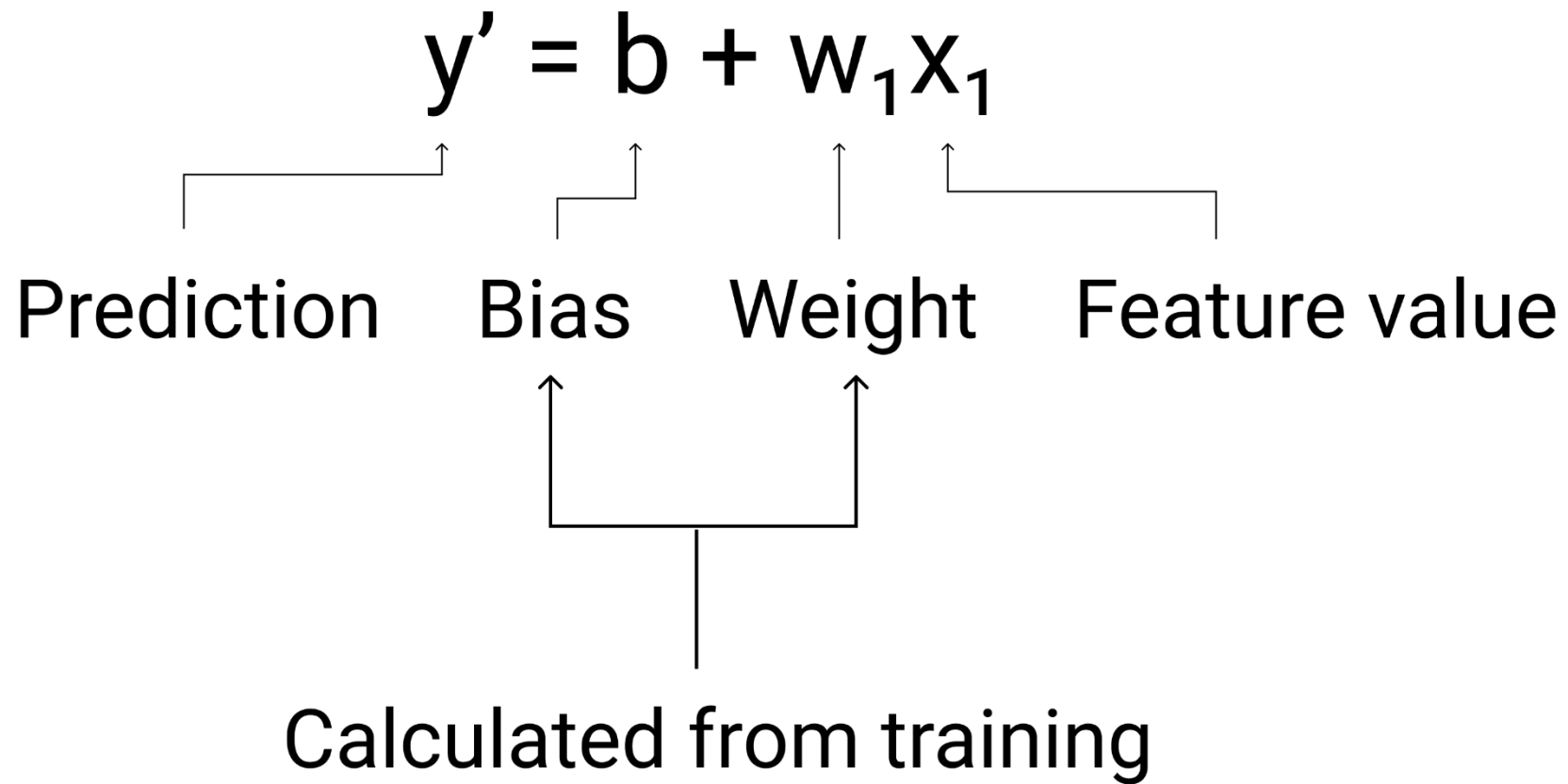
- $y$  é o valor em milhas por galão que queremos prever.
- $m$  é a inclinação da linha.
- $x$  é o valor de entrada em libras.
- $b$  é a interseção com o eixo  $y$ .

# Em aprendizado de máquina

$$y' = b + w_1 x_1$$

- ✓  $y'$  é o **rótulo previsto**, ou seja, a **saída**
- ✓  $b$  é o **viés (bias)** do modelo.
  - ✓ É o mesmo conceito da **interseção  $y$**  na equação algébrica. É um parâmetro do modelo que é calculado durante o treinamento.
- ✓  $w_1$  é o **peso** do elemento.
  - ✓ É o mesmo conceito da **inclinação  $m$**  na equação algébrica. É um parâmetro do modelo que também é calculado durante o treinamento.
- ✓  $x_1$  é um **atributo**, ou seja, a **entrada**.

# Durante o treinamento



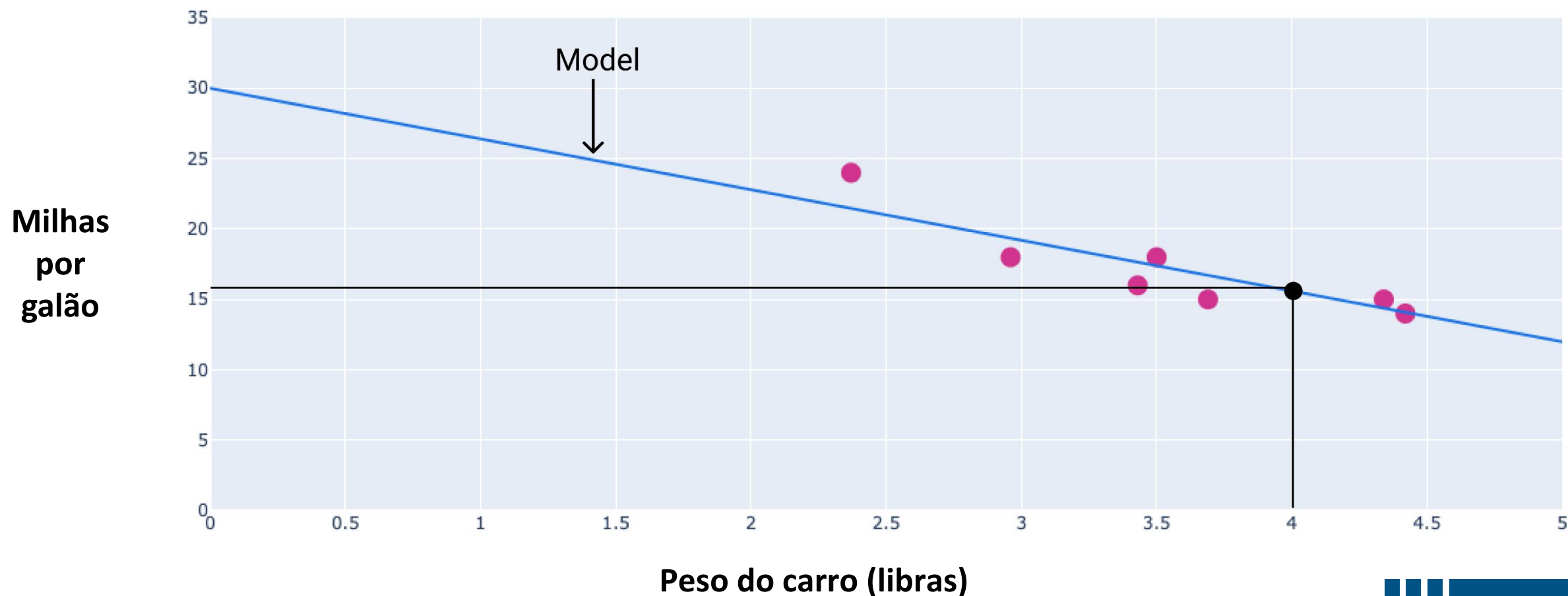


# Exemplo

O viés é 30 (onde a linha cruza o eixo y), e o peso é -3,6 (a inclinação da linha).

Com isso, teríamos:  $y' = 30 + (-3,6) * x_1$

Exemplo de previsão:  $y' = 30 + (-3,6) * 4.000 = 15,6$  milhas por galão



# Regressão múltipla

Um carro pode ser descrito por várias outras características  $x$ , cada qual com seu peso ( $w$ ).

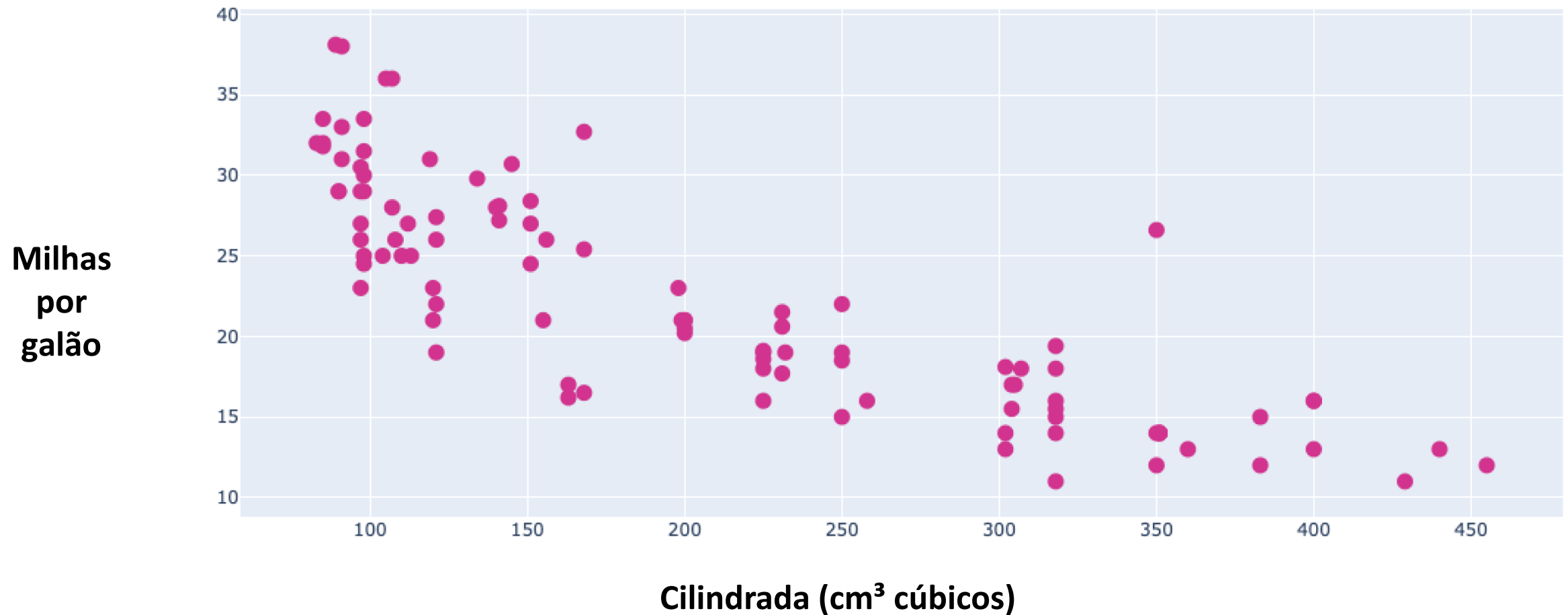
Milhas por galão

$$y' = b + w_1 X_1 + w_2 X_2 + w_3 X_3 + w_4 X_4 + w_5 X_5$$

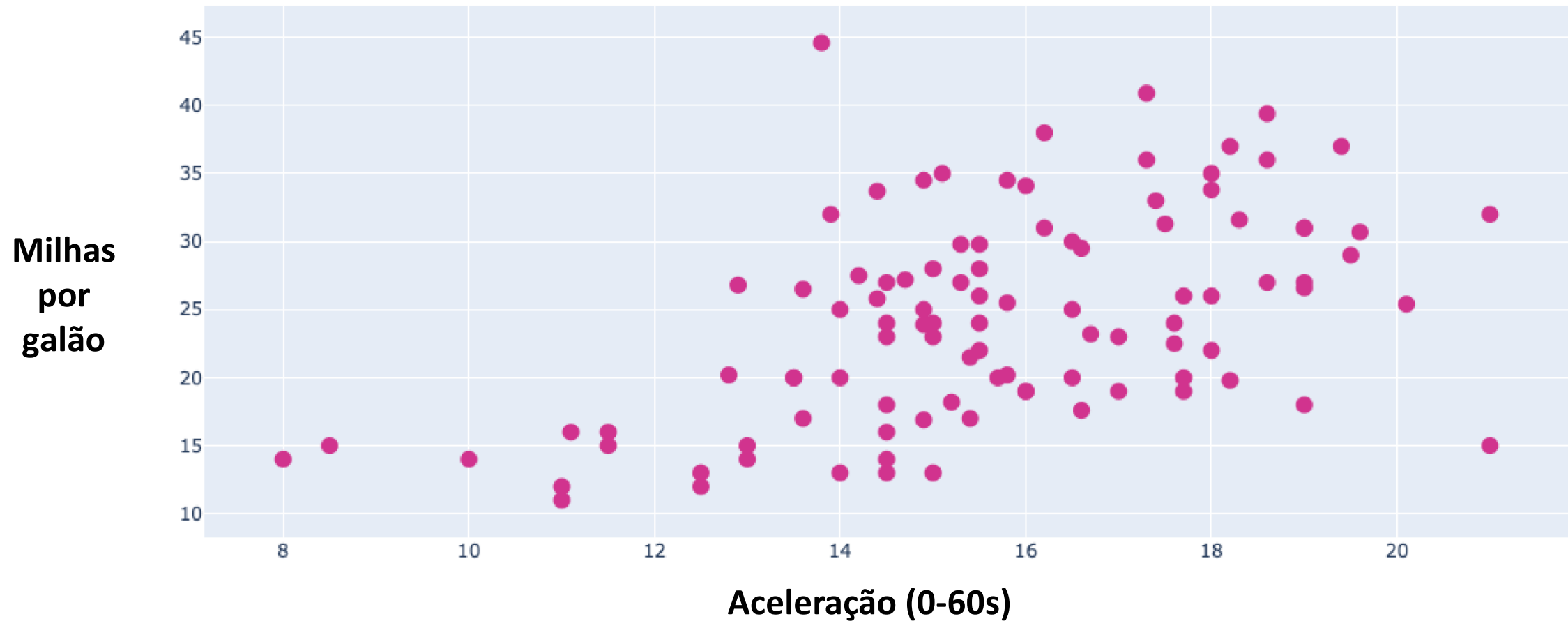
The diagram illustrates the relationship between car features and the regression equation. Arrows point from the feature names below to the corresponding terms in the equation above:

- Peso** points to  $w_1$
- Cilindrada** points to  $X_1$
- Aceleração** points to  $w_3$
- Nº de cilindros** points to  $X_4$
- Potência** points to  $X_5$

# Exemplo



# Exemplo



# Exemplo

