Aprendizado de Máquina

Aula 2.1 - Algoritmos de Regressão

Adriano Rivolli

rivolli@utfpr.edu.br

Especialização em Inteligência Artificial

Universidade Tecnológica Federal do Paraná (UTFPR)
Câmpus Cornélio Procópio
Departamento de Computação

Conteúdo

Introdução

Regressão

- Tarefa do aprendizado supervisionado
- O atributo alvo é um valor real
- É usada para prever o valor de uma **variável dependente** com base em uma ou mais **variáveis independentes**
- Explora o relacionamento entre as variáveis do problema

Aplicações

- Previsão de custos, preços e vendas
- Estimativa de consumo e demanda
- Previsão de pontuação, crédito e taxas
- Preenchimento de valores ausentes

Algoritmos

- Regressão Linear
- Regressão Polinomial
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression
- K-nearest neighbors regression

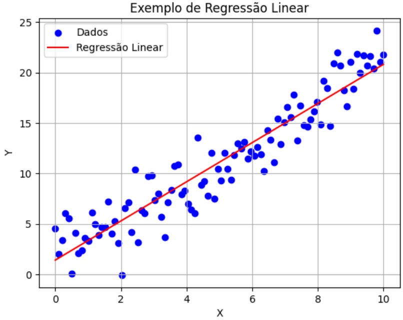
Regressão linear

Regressão Linear

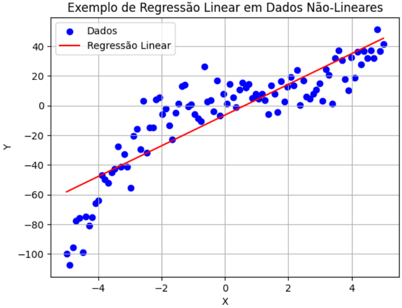
- Assume que os dados (entrada/saída) são lineares (**viés do algoritmo**)
- Aprende uma função linear que melhor aproxima a entrada em saída
- Há diferentes formas de estimar modelos lineares

Vamos conhecer o Mínimos Quadrados (*Least Squares*)

Exemplo



Exemplo de Regressão Linear
Dados
Regressão Linear
Y
X
0
5
10
15
20
25
0
2
4
6
8
10



Exemplo de Regressão Linear em Dados Não-Lineares
Dados
Regressão Linear
Y
X
−100
−80
−60
−40
−20
0
20
40
−4
−2
0
2
4

Modelo

$$\hat{y} = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$$

- β_0 intercepto (também chamado de viés)
- β_i coeficientes associados às variáveis preditivas
- x é a instância que se deseja prever o valor de $\hat{y}$

- O treinamento consiste em aprender os valores de β

Least Squares (exemplo)

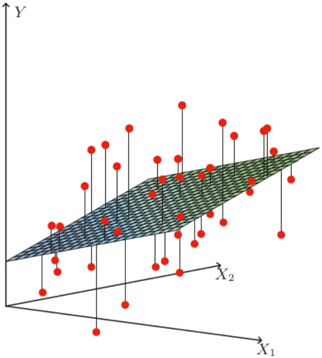
- os coeficientes β são definidos visando diminuir a soma dos erros residuais

  RSS = *Residual Sum-of-Squares*

$$\mathbf{RSS}(\beta) = \sum_{i=1}^{N}(y_i - \hat{y})^2$$

$$\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_i x_{ij})^2$$

Erros residuais



Fonte: Hastie, T., Tibshirani, R., Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.), pg. 45

Interpretando os coeficientes

- β_0 **intercepto**: Valor predito quando todas as variáveis independentes são iguais a zero
- **Coeficientes** β_i: Cada coeficiente está associado a uma variável independente específica
  - Coeficiente positivo sugere que um aumento na variável independente está associado a um aumento no valor predito
  - Coeficiente negativo é o contrário
  - Valores próximos a 0 indicam que a variável independente em questão não possui relevância para a predição

Erro padrão e Z-score

- Além dos valores dos coeficientes é possível entender a confiabilidade e significância estatística dos valores estimados

- ***Std. Error* (Erro Padrão)**: Quanto menor, mais confiável é a estimativa do coeficiente
- **Z-score**: Valor absoluto maior que 1.96 indica que o coeficiente é estatisticamente significativo ao nível de significância de 5%

Exemplo

Term	Coefficient	Std. Error	Z Score
`Intercept`	2.46	0.09	27.60
`lcavol`	0.68	0.13	5.37
`lweight`	0.26	0.10	2.75
`age`	−0.14	0.10	−1.40
`lbph`	0.21	0.10	2.06
`svi`	0.31	0.12	2.47
`lcp`	−0.29	0.15	−1.87
`gleason`	−0.02	0.15	−0.15
`pgg45`	0.27	0.15	1.74

Fonte: Hastie, T., Tibshirani, R., Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.), pg. 50

Regularização

Evitando overfitting

- A regularização é uma penalidade adicionada a função de custo
- Utilizada para controlar a magnitude dos coeficientes
- O seu uso faz com que o valor de alguns coeficientes sejam próximos de 0 (penaliza coeficientes grandes)
- Ajuda na seleção dos atributos mais relevantes

Lasso

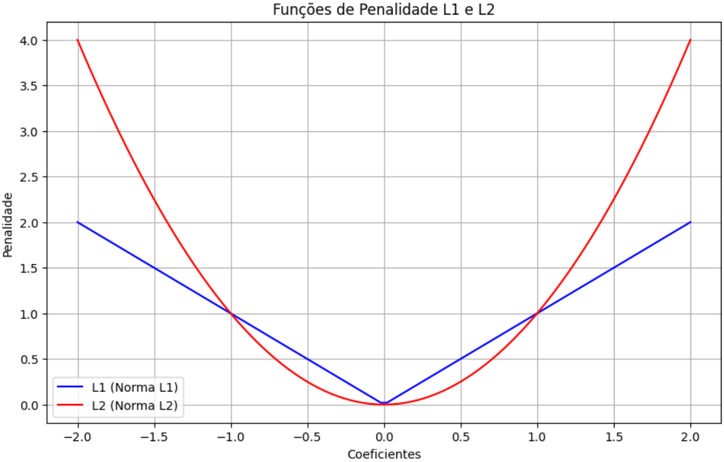
- *Least Absolute Shrinkage and Selection Operator*
- Útil para dados com alta dimensionalidade e atributos irrelevantes
- Utiliza a norma L1 (distância Manhattan)
- $Custo = \mathbf{RSS} + \alpha \sum_{i=1}^{p} |\beta_i|$
  - α controla a força da regularização

Ridge

- Útil para dados com atributos redundantes
- Utiliza a norma L2 (distância Euclidiana)

- $Custo = \textbf{RSS} + \alpha \sum_{i=1}^{p} \beta_i^2$
  - α controla a força da regularização

L1 vs L2


Funções de Penalidade L1 e L2
Penalidade
Coeficientes
L1 (Norma L1)
L2 (Norma L2)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

L1 vs L2 (características)

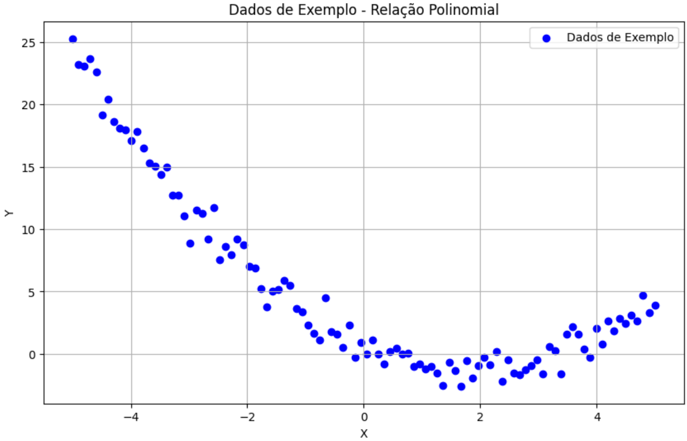
- **L1**: soma dos valores absolutos dos coeficientes
  - Tende a produzir coeficientes esparsos ($\beta_i = 0$)
  - Seleção de variáveis
- **L2**: soma dos quadrados dos coeficientes
  - Penalização suave e distribuída
  - Reduz a magnitude de todos os coeficientes proporcionalmente

Regressão Polinomial

Transformação dos dados

- Um caso especial de regressão linear
- Aplica-se um mapeamento dos dados
- O modelo se torna:
  $\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n$
  - n é grau do polinômio
- É possível capturar relacionamentos mais complexos nos dados

Dados polinomiais



Atributos polinomiais

- Dado 2 atributos $[a, b]$
- Aplicação da transformação polinomial de grau 2 resulta:
  - $[1, a, b, a^2, ab, b^2]$
- O número de atributos gerados é dado por:

$$\frac{(p+d)!}{p!d!},$$

onde p é o grau do polinômio e d o número de atributos

A escolha do grau