



Regularização em Regressões

☰ Ciclo	Ciclo 07: Outros algoritmos Classificação e Regressão
# Aula	55
🕒 Created	@April 13, 2023 1:32 PM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ A regularização L1 (Lasso)
- ☐ A regularização L2 (Ridge)
- ☐ A regularização L1 e L2 (Elastic Net)
- ☐ Quando usar cada regularização?
- ☐ Resumo
- ☐ Próxima aula

Conteúdo:

▼ 1. A regularização L1 (Lasso)

A regularização L1 é uma técnica utilizada na Regressão Logística para evitar o overfitting, adicionando uma penalidade à função de custo para evitar que os parâmetros do modelo sejam muito grandes.

$$J(\theta) = \frac{-1}{m} \sum [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))] + \lambda \sum |\theta_j|$$

onde:

- lambda é o parâmetro de regularização que controla a força da penalidade
- |theta| é o valor absoluto do j-ésimo parâmetro do modelo.

A adição da penalidade de regularização L1 incentiva o modelo a utilizar apenas os parâmetros mais importantes, zerando os parâmetros menos relevantes.

▼ 2. A regularização L2 (Ridge)

A penalidade L2 é proporcional ao quadrado da magnitude dos parâmetros, incentivando o modelo a ter parâmetros menores.

$$J(\theta) = \frac{-1}{m} \sum [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))] + \frac{\lambda}{2} \sum (\theta_j)^2$$

onde:

- lambda é o parâmetro de regularização que controla a força da penalidade.
- $(\theta_j)^2$ é o valor absoluto do j-ésimo parâmetro do modelo.

A penalidade L2 incentiva o modelo a ter parâmetros menores, pois a função de custo é minimizada quando os valores dos parâmetros são pequenos.

A regularização L2 também tem uma propriedade chamada "smoothing" (suavização), que suaviza as diferenças entre os valores dos parâmetros, reduzindo a complexidade do modelo e evitando o overfitting.

▼ 3. A regularização L1 e L2 (Elastic Net)

A penalidade L2 é proporcional ao quadrado da magnitude dos parâmetros, incentivando o modelo a ter parâmetros menores.

A regularização Elastic Net é útil quando há muitas características nos dados que podem ser irrelevantes para a predição.

O termo de regularização L1 ajuda a tornar os coeficientes dessas características irrelevantes iguais a zero, enquanto o termo L2 ajuda a evitar a superestimação dos coeficientes restantes.

$$J(\theta) = \frac{-1}{m} \sum [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))] + \lambda_1 \sum |\theta_j| + \lambda_2 \sum (\theta_j)^2$$

onde:

- lambda 1 é o parâmetro de regularização que controla a força da penalidade L1
- lambda 2 é o parâmetro de regularização que controla a força da penalidade L2

1. Quando λ_1 é zero, a regularização é equivalente à regularização L2 (Ridge).
2. Quando λ_2 é zero, a regularização é equivalente à regularização L1 (Lasso).
3. Quando λ_1 e λ_2 estão entre 0 e 1, a regularização é uma combinação ponderada das duas.

▼ 4. Quando usar cada regularização?

▼ 4.1 Quando usar a regularização L1 ?

Quando não houver uma seleção de features antes do treinamento do algoritmo

A regularização L1 é particularmente útil quando o conjunto de dados de treinamento tem muitas características, algumas das quais podem ser irrelevantes ou redundantes. A regularização L1 pode ajudar a selecionar as características mais relevantes para a predição, eliminando as características irrelevantes.

▼ 4.2 Quando usar a regularização L2 ?

Quando tiver ocorrido uma seleção de features antes do treinamento do algoritmo

Em geral, a regularização L2 é útil quando o modelo tem muitos recursos e o conjunto de dados de treinamento é relativamente pequeno. Ela também é eficaz quando os recursos são altamente correlacionados, pois ajuda a reduzir a multicolinearidade.

▼ 4.3 Quando usar a regularização Elastic Net ?

Quando tiver ocorrido uma seleção de features antes do treinamento do algoritmo e você quer usar uma segunda seleção de features.

A regularização Elastic Net é uma combinação da regularização L1 e L2 e permite controlar a força relativa de cada técnica de regularização usando um parâmetro de elasticidade. Isso pode torná-la mais flexível do que as outras duas técnicas, mas também mais difícil de ajustar.

▼ 5. Resumo

1. Use a regularização L1 (Lasso) se você acredita que muitas das características do seu conjunto de dados são irrelevantes e pode ser benéfico eliminá-las do modelo. A regularização

L1 tem o efeito de definir muitos coeficientes do modelo para zero, o que pode tornar o modelo mais fácil de interpretar. Ela é particularmente útil quando há muitas características no conjunto de dados e você deseja reduzir a dimensionalidade do modelo.

2. Use a regularização L2 (Ridge) se você acredita que todas as características do seu conjunto de dados são relevantes para a predição e deseja evitar que os coeficientes do modelo fiquem muito grandes. A regularização L2 tem o efeito de reduzir o tamanho geral dos coeficientes do modelo, mas não os define como zero. Ela é particularmente útil quando você tem muitas características e não deseja descartar nenhuma delas.
3. Use a regularização Elastic Net se você acredita que algumas características do seu conjunto de dados são irrelevantes e outras são importantes para a predição.

▼ 6. Próxima aula

Regularização na prática