Regressão Ridge e Lasso

Modelos Preditivos

Grande parte dos modelos elaborados na mineração de dados tem por objetivo fazer predição.

Isso só é possível se tivermos certeza de que os dados mantém o mesmo padrão através da janela de tempo considerada.

Essa propriedade é chamada estacionalidade.

Modelos Preditivos

Um modelo preditivo pode ser usado para:

Obter uma regra com o objetivo de

- classificar,
- ordenar ou
- estimar novos casos;

Destacar variáveis exploratórias úteis dentre as inúmeras disponíveis;

Otimizar a complexidade do processo com o objetivo de reduzir as perturbações aleatórias das observações de treinamento.

Modelos Preditivos

Uma forma empírica e muito utilizada de verificar o poder de predição de um modelo é particionar as observações em pelo menos dois conjuntos treinamento e validação

A partir do conjunto de treinamento serão gerados os valores preditos pelo modelo.

O conjunto de validação é usado para avaliar o modelo ajustado na forma de estatísticas numéricas e gráficas.

Problemas que um modelo pode apresentar:

- não ser flexível o suficiente para captar sinais importantes (underfitting ou alto bias).

- ser tão complexo que acompanha as variações aleatórias para uma particular amostra (overfitting ou alta variância).

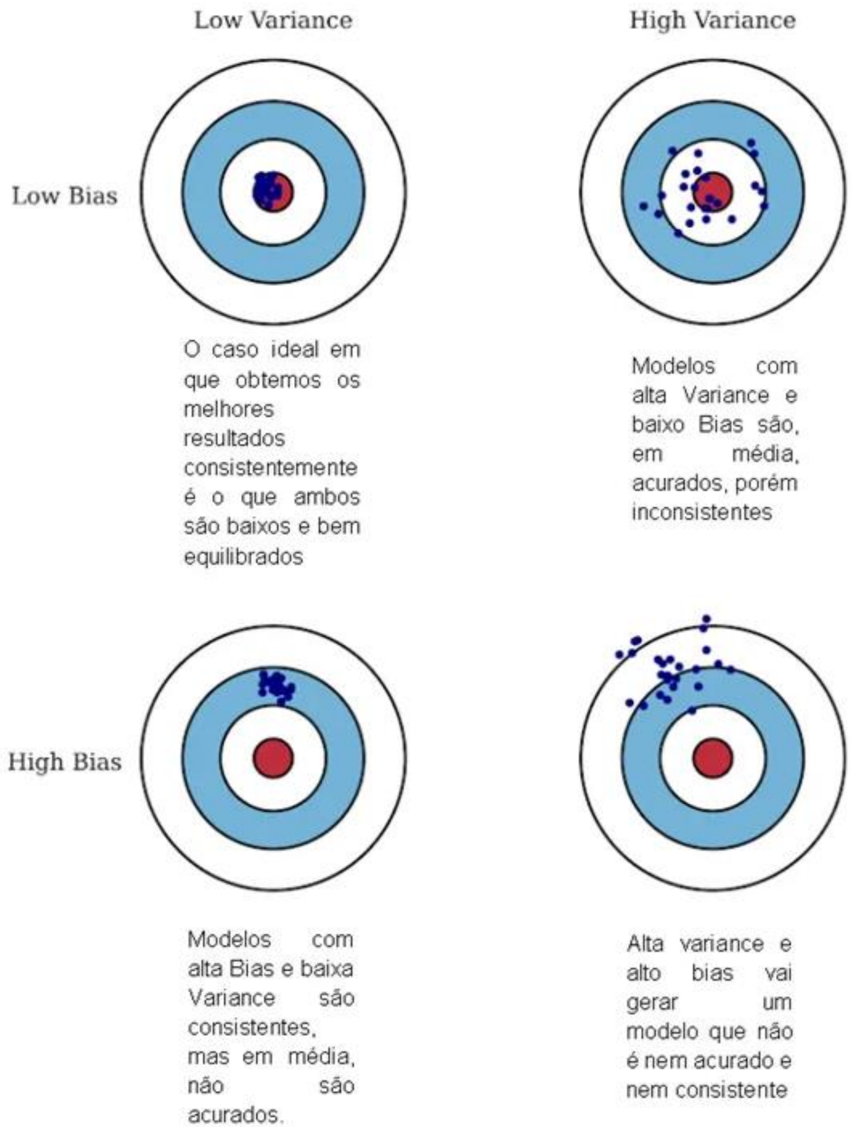
Bias e Variance

Para entender melhor os erros de predição é necessário entender que fatores influenciam tais erros, bem como a relação entre eles.

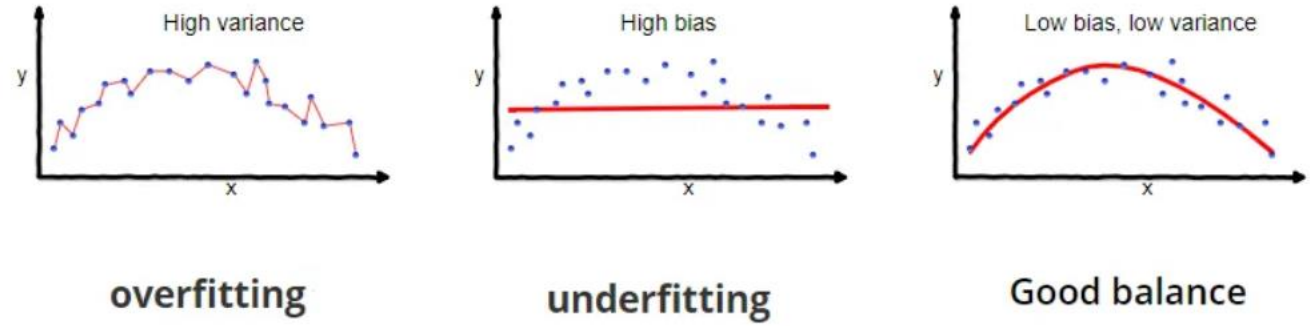
O primeiro fator, chamado de bias (viés), é a diferença entre a predição média do nosso modelo e o valor correto esperado. Sendo assim, um modelo com bias aprende relações erradas e gera previsões longe do esperado. Dessa forma, modelos com alto bias possuem um problema de underfitting.

O segundo fator, variance, é a variabilidade do modelo para um determinado dado, ou seja, a capacidade do modelo de se adaptar à base de treino e ao ruído. Dessa forma, modelos com alta variance focam excessivamente se ajustar aos dados e, inclusive, ao ruído. Assim, esses modelos têm um problema de overfitting, ou seja, se adaptam tão bem ao dataset que não conseguem generalizar para além dele.

Bias e Variance



Bias e Variance



Relembrando – Regressão Linear Múltipla

- Reta Estimada

$$Y_i = a + b_1 . X_{1i} + b_2 . X_{2i} + ... + b_k . X_{ki} + e_i$$

a - COEFIC. LINEAR DA FÓRMULA
Valor do eixo vertical interceptado pela reta (Intercepto do Eixo y).

b_i - COEFIC. ANGULAR DA i-ésima VARIÁVEL

i – Número de Variáveis Independentes

- **Método dos Mínimos Quadrados:** Obter a reta que melhor se ajusta aos dados. Busca reduzir a soma do quadrado dos resíduos

$$RSS = \Sigma(y_i - \hat{y}_i)^2$$

Regularização

A regularização é uma técnica utilizada nos modelos de Ridge e Lasso para evitar o sobreajuste (overfitting) dos modelos de Regressão Linear, especialmente em situações onde há multicolinearidade entre as variáveis independentes.

Nos dois casos, a regularização envolve adicionar um termo de penalização à função de custo do modelo, que depende dos coeficientes estimados. Essa penalização tem o efeito de restringir os valores dos coeficientes, limitando sua magnitude. Isso ajuda a controlar a complexidade do modelo, evitando que os coeficientes se tornem muito grandes em magnitude.

Regressão de Ridge

- A regressão Ridge adiciona um termo de penalidade à função de custo do modelo linear, que é proporcional ao quadrado dos coeficientes de regressão.

- Isso significa que, durante o treinamento, a regressão Ridge tenta minimizar a soma dos quadrados dos coeficientes, além da soma dos quadrados dos resíduos.

- A penalização adicionada pela regressão Ridge ajuda a reduzir a variância dos estimadores, o que pode melhorar a generalização do modelo.

- Uma característica importante da regressão Ridge é que ela não exclui variáveis do modelo, mas apenas as reduz em importância, tornando-as mais robustas a variações nos dados.

Regressão de Ridge

- Na Regressão Linear, busca-se reduzir Busca reduzir a soma do quadrado dos resíduos

$$RSS = \Sigma(y\, i - \hat{y}\, i\,)^2$$

- A Regressão de Ridge introduz um termo de regularização chamado de L2.

$$\lambda\, \Sigma\, b_j{}^2$$

- Nesse caso busca-se reduzir

$$RSS + \lambda\, \Sigma\, b_j{}^2$$

- Quando $\lambda = 0$, este termo de penalidade não tem efeito e a regressão de crista produz as mesmas estimativas de coeficiente que os mínimos quadrados.

Regressão de Ridge

Passos para executar uma Regressão de Ridge

1. Padronize cada variável preditora.

2. Ajuste o modelo de regressão de Ridge e escolha um valor para λ

Não existe uma fórmula exata que possamos usar para determinar qual valor usar para λ. Na prática, existem duas maneiras comuns de escolher λ:

   1. Crie um gráfico de rastreamento Ridge.
   2. Calcule o erro quadrático médio(MSE) para cada valor de λ.

Regressão Lasso (Least Absolute Shrinkage and Selection Operator):

- Assim como a Ridge, a regressão Lasso também adiciona um termo de penalidade à função de custo, mas esse termo é proporcional ao valor absoluto dos coeficientes de regressão.

- Uma das principais diferenças entre a Lasso e a Ridge é que a Lasso pode levar alguns coeficientes de regressão a zero, efetivamente fazendo uma seleção de variáveis.

- Isso significa que a regressão Lasso não apenas reduz a variância, mas também pode realizar uma seleção automática de variáveis, excluindo aquelas menos importantes para a previsão.

- A seleção automática de variáveis torna a regressão Lasso útil em situações em que há muitas variáveis preditoras, algumas das quais podem não contribuir significativamente para a previsão.

Regressão LASSO (Least Absolute Shrinkage and Selection Operator):

- Na Regressão Linear, busca-se reduzir Busca reduzir a soma do quadrado dos resíduos

$$RSS = \Sigma(y\ i - \hat{y}\ i\ )^2$$

- A Regressão LASSO introduz um termo de regularização chamado de L1.

$$\lambda \Sigma\ |b_j|$$

- Nesse caso busca-se reduzir

$$RSS + \lambda \Sigma\ |b_j|$$

- Quando $\lambda = 0$, este termo de penalidade não tem efeito e a regressão de crista produz as mesmas estimativas de coeficiente que os mínimos quadrados.

Regressão LASSO

Passos para executar uma Regressão LASSO

1. Padronize cada variável preditora.

2. Ajuste o modelo de regressão LASSO e escolha um valor para λ

Não existe uma fórmula exata que possamos usar para determinar qual valor usar para λ. Na prática, existem duas maneiras comuns de escolher λ:

   1. Crie um gráfico de rastreamento LASSO.
   2. Calcule o erro quadrático médio(MSE) para cada valor de λ.

Regressão Elastic Net

Elastic Net é uma técnica de regularização que combina os aspectos da regressão Ridge (L2 regularization) e da regressão Lasso (L1 regularization). Ele é útil quando há multicolinearidade entre os preditores (variáveis independentes), o que pode causar instabilidade ou alta variância nos coeficientes estimados.

Ao combinar as penalidades L1 e L2, o Elastic Net consegue superar algumas das limitações de cada uma dessas técnicas individualmente. Ele utiliza dois parâmetros de regularização: alpha (α) e lambda (λ).

$$RSS + \lambda \left(\alpha \Sigma |b_j| + (1-\alpha) \Sigma b_j^2\right)$$

Regressão Elastic Net

Passos para executar uma Elastic Net

1. Padronize cada variável preditora.

2. Ajuste o modelo de regressão LASSO e escolha um valor para λ e um para α

Não existe uma fórmula exata que possamos usar para determinar qual valor usar para λ e α. A mesma lógica anterior é aplicada