

# Regularização para Regressão

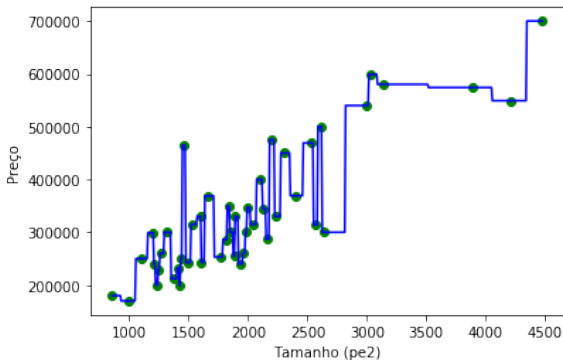
---

Advanced Institute for Artificial Intelligence

<https://advancedinstitute.ai>

# Regressão: Onde estamos

Mesmo com modelos relativamente simples, conseguimos fazer um fit bom no conjunto de treinamento



## Seria mesmo?

- Modelos complexos tem tendência a se especializarem demais ao conjunto de treinamento.
- Modelos que aderem perfeita e completamente aos exemplos mostrados correm risco de *overfitting*

## Mas o que teria de ruim em aderir perfeitamente aos dados?

- **Lembre-se:** O Conjunto de treinamento não é uma representação perfeita do mundo real
- O objetivo principal é modelar um fenômeno através de exemplos, saber classificar apenas o conjunto de treinamento não vale de nada

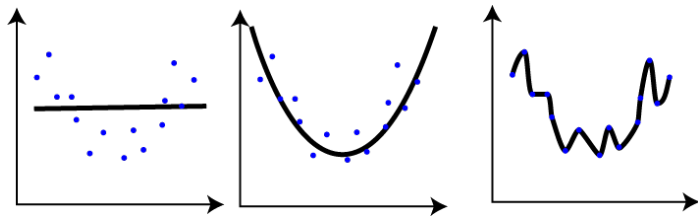
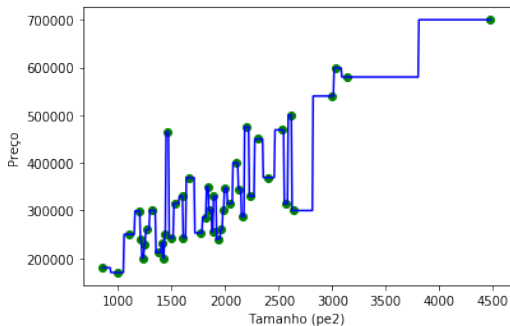
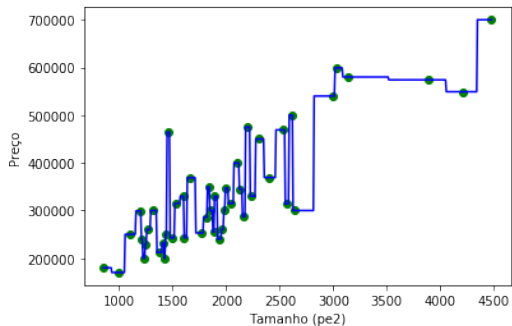


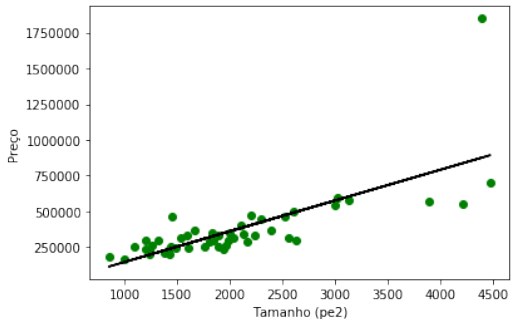
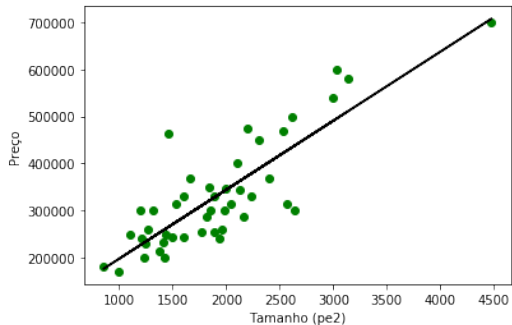
Figure: (Esquerda): Underfit, (Centro): Fit, (Direita): Overfit

# Visualizando Overfitting

## Como o modelo da árvore muda ao retirar alguns exemplos



## Modelo linear tem sensibilidade a outliers



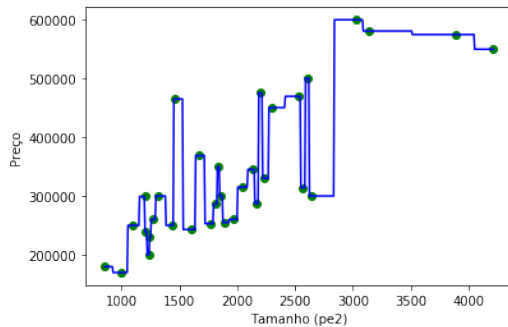
# Como saber se o modelo tem overfitting?

- Sempre avaliar modelos em amostras que o **modelo nunca viu**
- Fazer divisões em bases de treinamento e teste, possivelmente utilizando validação cruzada

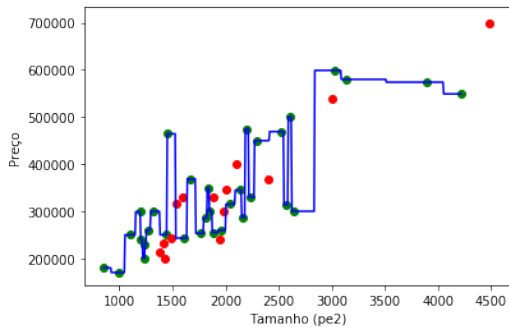


# Estudo base Portland: Árvore

Base de treinamento:  $R^2 = 1.0$



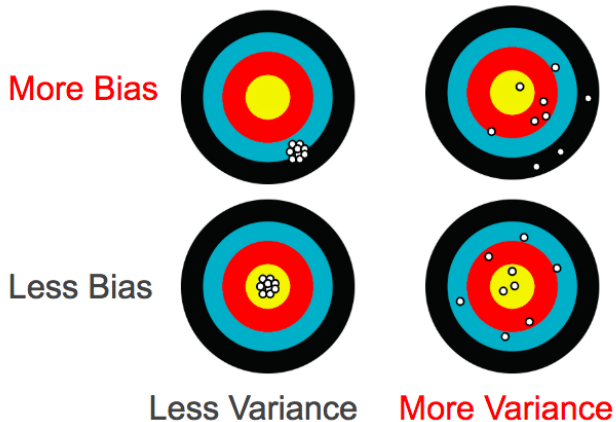
Base de teste:  $R^2 = 0.43$



## Métrica perfeita na base de teste significa que o modelo é perfeito?

- **Não necessariamente:** Em nenhum problema não-trivial você terá acesso à uma base de dados completamente representativa do problema
- Avaliar com uma base de teste *alivia*, não resolve o problema
- Nunca haverá exemplos suficientes para modelar perfeitamente o fenômeno

# Dilema: Variância x Viés



- Bias é a diferença entre a previsão média do modelo e o valor correto que estamos tentando prever. Bias alto significa que o modelo é simplificado demais para os dados.
- Variância é a variabilidade da previsão do modelo para um dado ponto de dados. O modelo com alta variação presta muita atenção aos dados de treinamento e não generaliza suficientemente para os dados que não conhece.

- Modelo muito simples com poucos parâmetros tem Bias alto e baixa variação.
- Modelo complexo com um grande número de parâmetros, terá alta variação e Bias baixo.
- Deve-se buscar o equilíbrio / bom sem sobreajustar e não adequar os dados.

- Os modelos devem tentar **generalizar** além do que é observado no conjunto de treinamento.
- A **Regularização** tem o papel de controlar o overfitting dos classificadores.

- Diminui a **variância** reduzindo a eficácia no aprendizado
- Penaliza a complexidade do modelo
- Quase todos os algoritmos de aprendizado possuem algum mecanismo de regularização



## Regularização L2

Busca diminuir o valor dos parâmetros da função de regressão

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^n (y_i - (\theta_0 + \theta_1 x_i))^2 + \sum_{\theta_i} \frac{\lambda}{2n} \theta_i^2$$

## Regularização L1

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^n (y_i - (\theta_0 + \theta_1 x_i))^2 + \sum_{\theta_i} \frac{\lambda}{2n} \theta_i$$

Dois exemplos de procedimentos de regularização para regressão linear são:

- Regressão lasso: mínimos quadrados são modificados para também minimizar a soma absoluta dos coeficientes (regularização L1).
- Regressão Ridge: Mínimos Quadrados são modificados para também minimizar a soma absoluta ao quadrado dos coeficientes (chamada regularização L2).

- Ao utilização de um modelo de regularização leva a uma normalização do modelo.
- Regressão Ridge é útil quando todas as variáveis são relevantes.
- Regressão Lasso dá mais prioridade as variáveis que impactam mais no modelo, é recomendável para casos onde a influência das variáveis não é conhecida

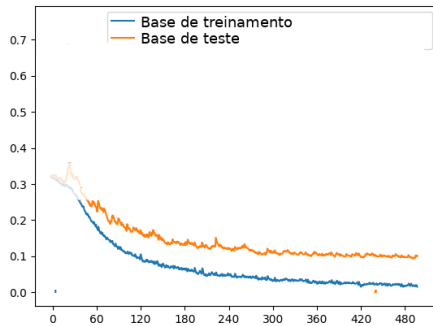
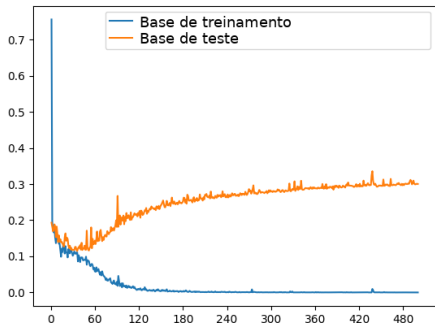
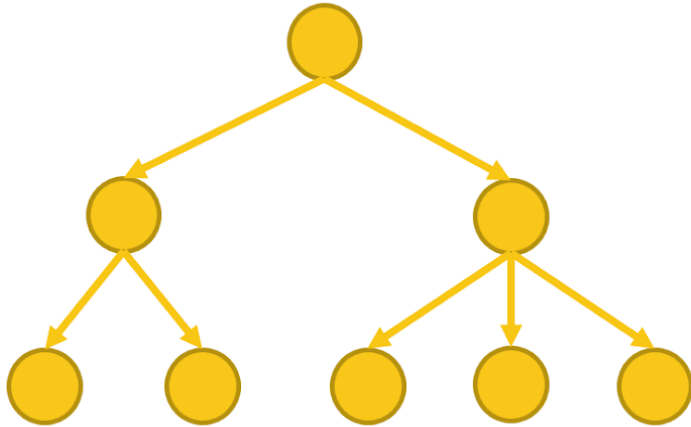


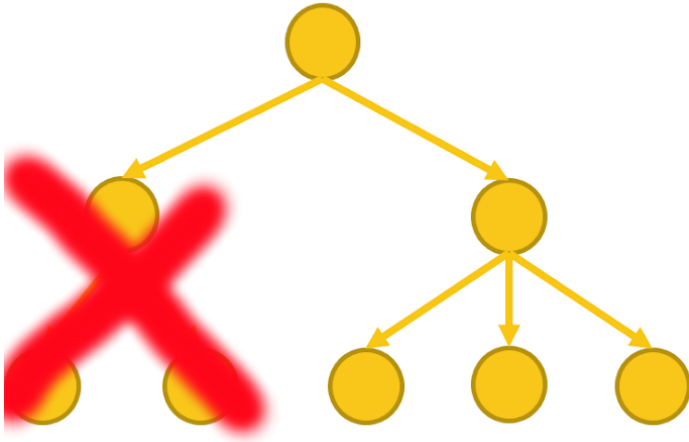
Figure: (esq): Overfitting (dir): Modelo geral

- De certa forma, o parâmetro  $K$  já funciona como um regularizador
- Quanto maior o  $K$ , mais "uniforme" a saída do regressor será

## A regularização da árvore é chamada Poda

- A ideia principal é "podar" algumas partes da árvore que estão atrapalhando na generalização







## A Poda pode ser feita de muitas maneiras:

- Limitando a altura da árvore
- Interrompendo o crescimento da árvore após um número de nós
- Eliminando nós que não representem um certo número de exemplos

- A regularização limita a complexidade dos modelos
- Seu objetivo é dar um peso à generalização, em contraponto à otimizar a métrica de desempenho
- Essencial para que os modelos sejam capazes de generalizar para exemplos não vistos