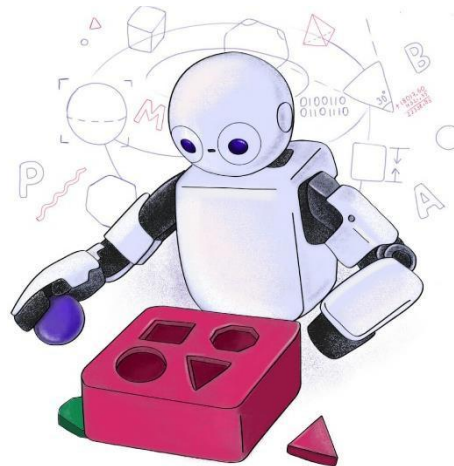


TP555 - Inteligência Artificial e Machine Learning: *Curvas de aprendizado*



Esse material foi desenvolvido e gentilmente cedido pelo Prof. Dr. Felipe Augusto Pereira de Figueiredo, do Inatel. (felipe.figueiredo@inatel.br)

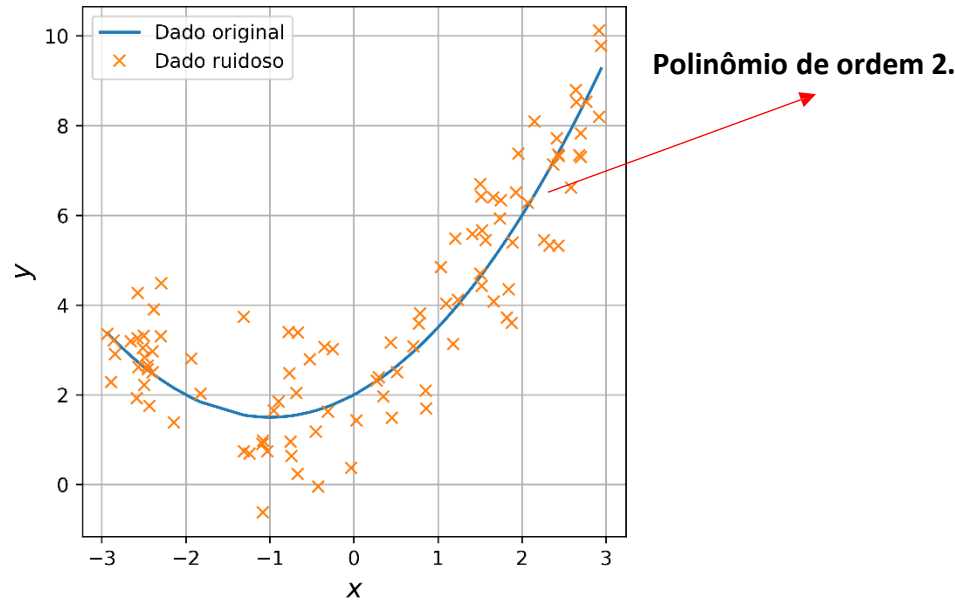
Inatel

Prof. Dr. Luiz Augusto Melo Pereira
luiz.melo@inatel.br

Curvas de Aprendizado

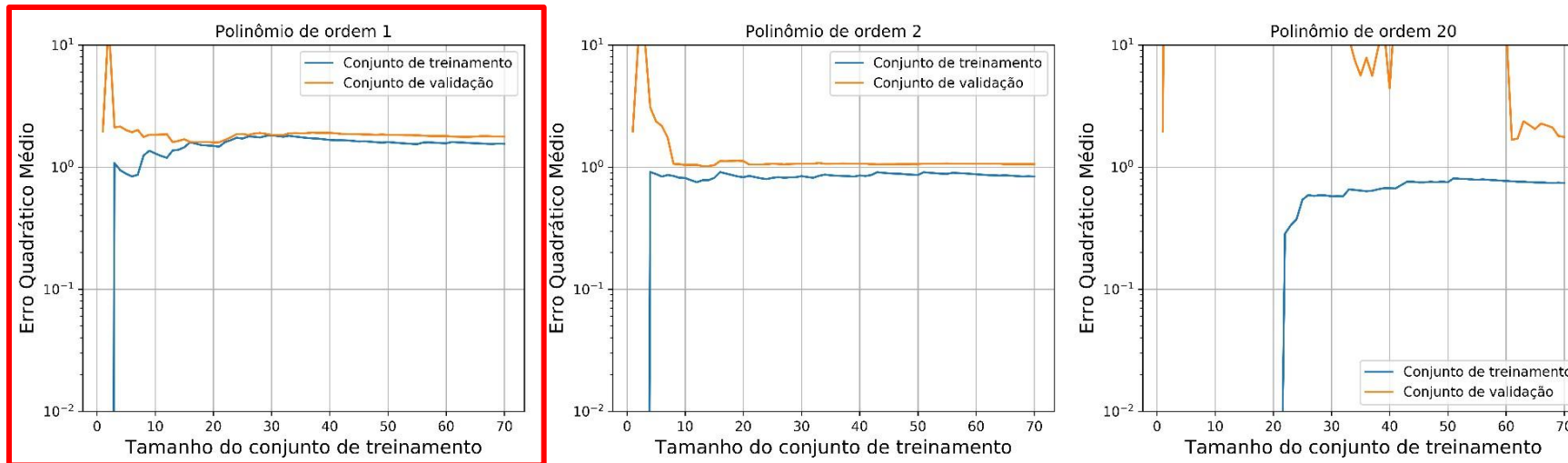
- **Curvas de aprendizado:** são gráficos que **comparam o desempenho do modelo nos conjuntos de treinamento e de validação em função do tamanho do conjunto de treinamento**.
 - **OBS.:** O tamanho do conjunto de validação permanece constante.
- São ferramentas úteis para **avaliar o desempenho do modelo de aprendizado de máquina à medida que mais dados são fornecidos para o treinamento**.
- Essa comparação é normalmente usada para avaliar:
 - Qual é o **melhor nível de complexidade de um modelo** (no caso de polinômios, sua ordem). Em outras palavras, são usadas para determinar o ponto de equilíbrio entre os erros de variância (sobreajuste) e de viés (subajuste).
 - O quanto o **modelo se beneficia de mais dados** (por exemplo, se temos "dados suficientes" ou se o desempenho do modelo ficará melhor se aumentarmos a base de dados).
 - Podemos também usar as curvas para **otimização dos hiperparâmetros** do modelo.

Curvas de Aprendizado: Exemplo



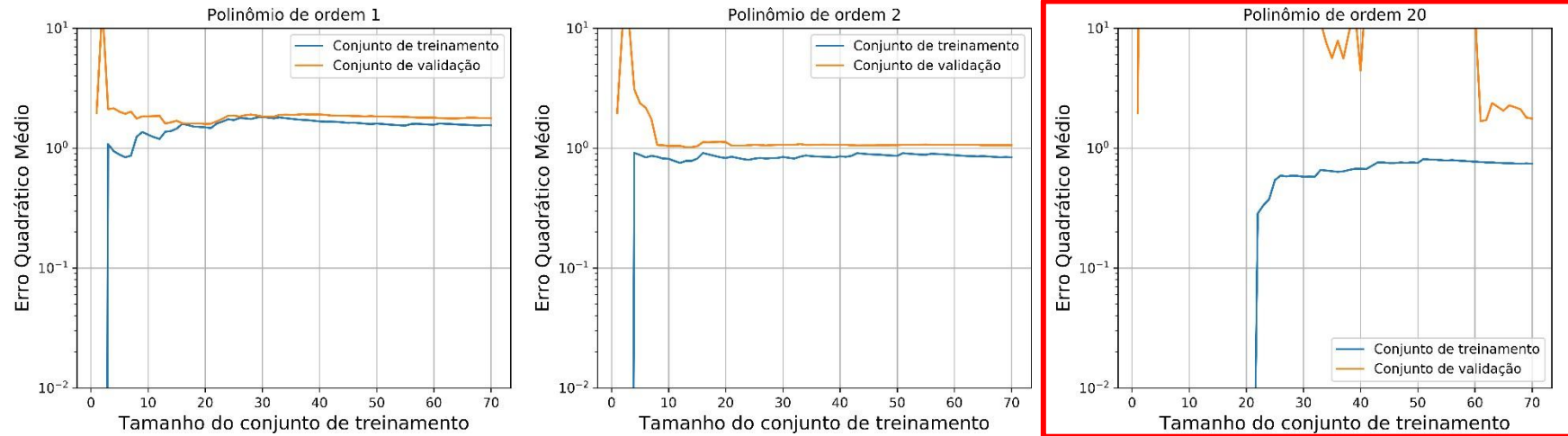
- Caso não conhecêssemos o ***mapeamento verdadeiro*** dos dados ruidosos da figura acima, qual ordem de polinômio melhor aproximaria o ***mapeamento***?
 - Além da validação cruzada, podemos usar as chamadas curvas de aprendizado para encontrar a ordem do polinômio aproximador.

Curvas de Aprendizado: *Polinômio de ordem 1*



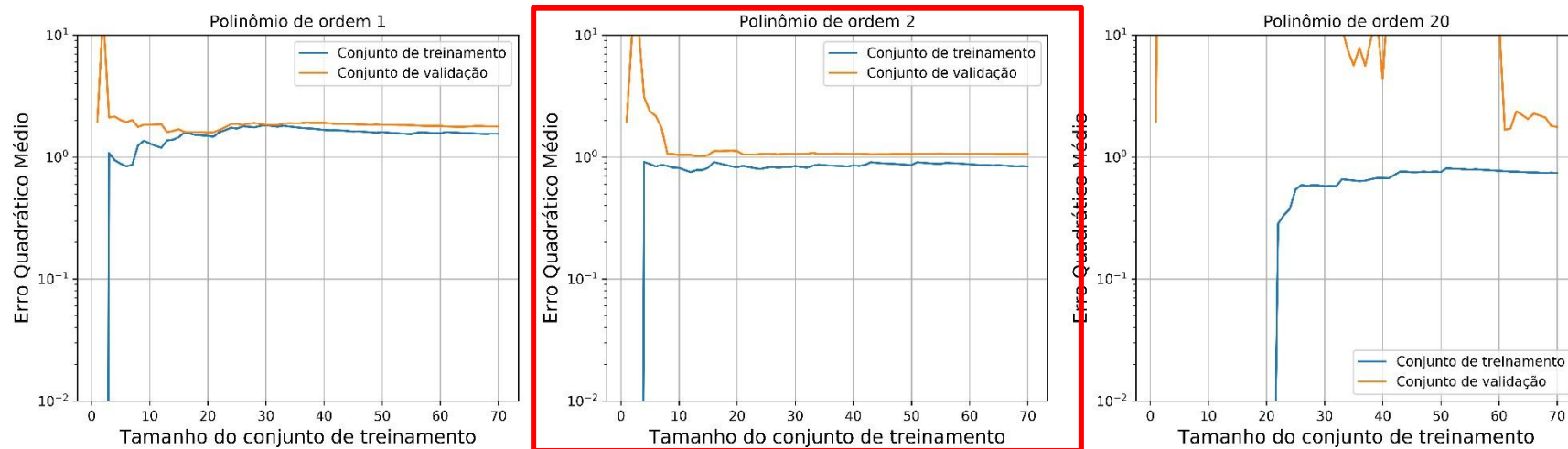
- Com 1 ou 2 exemplos no conjunto de treinamento, o modelo se ajusta perfeitamente (reta), porém, conforme o número de exemplos aumenta, é impossível para ele se ajustar (dados não-lineares e ruído).
- ***Erro nos exemplos de treinamento aumenta até atingir um platô e não diminui mesmo com o aumento do conjunto de treinamento***, pois o modelo não tem flexibilidade.
- ***Erro de validação é alto quando o modelo é treinado com poucos exemplos, porém, diminui conforme o conjunto de treinamento aumenta***, terminando em um platô próximo do erro de treinamento.
- Essas curvas são típicas de um modelo que está ***subajustando*** (erros altos e próximos).
- **O que fazer?** Aumentar a ordem do modelo.

Curvas de Aprendizado: *Polinômio de ordem 20*



- O erro de treinamento é menor do que com o modelo com polinômio de ordem 1.
- Porém, há uma diferença considerável entre as curvas do erro de treinamento e validação.
- Isso significa que o modelo tem um desempenho melhor no conjunto de treinamento do que no conjunto de validação, indicando que ele está **sobreaajustando**.
- A performance do modelo melhora caso o conjunto de treinamento aumente.
 - Com um conjunto maior, a tendência é que ambos os erros convirjam para o MSE mínimo.
- **O que fazer?** Diminuir a ordem do modelo ou aumentar o conjunto de treinamento.

Curvas de Aprendizado: *Polinômio de ordem 2*



- A diferença entre os erros diminui com o aumento do conjunto de treinamento se tornando pequena.
- Tanto o erro de treinamento quanto o de validação são menores do que com o modelo com polinômio de ordem 1 (i.e., reta).
- Isso é a indicação de um modelo que está se ajustando bem aos dados de treinamento (i.e., flexibilidade) e é capaz de generalizar bem para os dados de validação.
- Aumentar o conjunto de treinamento faz com que a diferença entre as duas curvas se torne ainda menor.
- **O que fazer?** Escolher esta ordem de polinômio.

Referências

1“learning-curve”, https://scikit-learn.org/stable/modules/learning_curve.html#learning-curve

2“what-is-a-learning-curve-in-machine-learning”,
<https://stackoverflow.com/questions/4617365/what-is-a-learning-curve-in-machine-learning>