



Lista de Exercícios 1

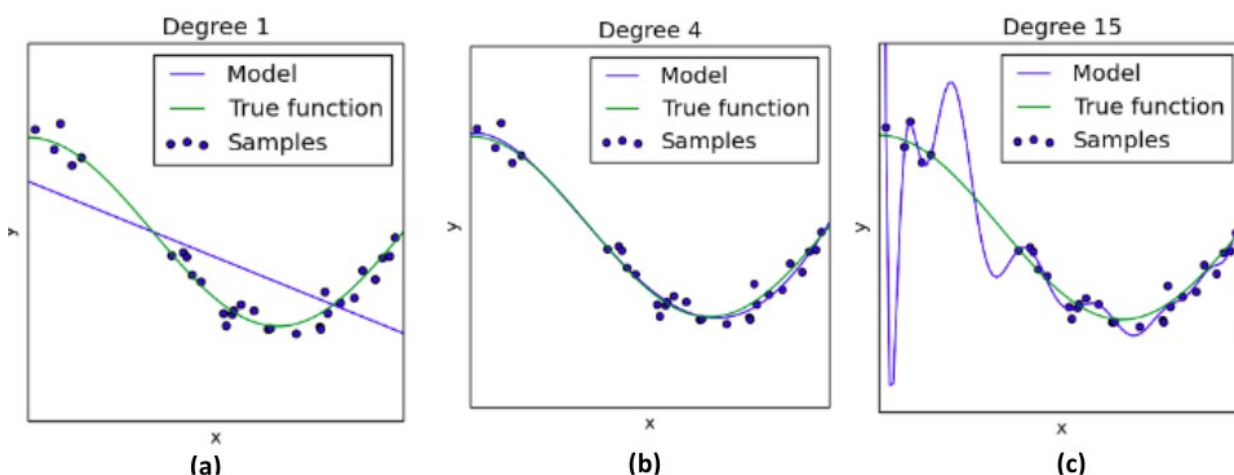
Entrega: 07/05

1. Dados um conjunto de exemplos de treinamento $X = \{x_1, x_2, \dots, x_n\}$ e um conjunto de classes $C = \{C_1, C_2, \dots, C_k\}$, um problema de classificação pode ser dividido em duas etapas: fase de inferência e fase de decisão.

Na fase de inferência, os dados de treinamento são utilizados para a definição de probabilidades a posteriori $p(C_k | x)$ e, na fase de decisão, os valores de probabilidade a posteriori são utilizados para fazer atribuições de classes para instâncias. Diante desse contexto, essas duas fases podem ser aprendidas por meio de duas abordagens:

- (1) separadamente via método bayesiano; ou
 - (2) conjuntamente por meio de aprendizagem de uma função que mapeia as instâncias de entrada diretamente às classes do problema, ou seja, a decisões.
- a) Explique com suas palavras como cada uma dessas duas abordagens funciona.
 - b) Descreva vantagens e desvantagens de cada abordagem

Para as questões a seguir, considere o exemplo de aproximação de função ilustrado nas imagens abaixo, as quais foram obtidas ao variar-se o grau do polinômio da função de aproximação.



2. Qual a relação entre o valor do grau do polinômio e os conceitos de *overfitting* e *underfitting*?



3. O ajuste do grau do polinômio e a busca pela função de aproximação mais adequada estão relacionados ao teorema *No Free Lunch*?
4. Esse algoritmo tenta minimizar o Erro Empírico? Por quê?
5. Sobre erro de generalização, responda:
 - a) Qual dos três modelos finais apresentará menor erro de generalização? Por quê?
 - b) Qual a relação entre erro de generalização e capacidade do modelo?
 - c) Qual a relação entre erro de generalização e dilema viés-variância?

Nos próximos exercícios, considere a seguinte relação:

A1	A2	A3	A4	Classe
F	F	V	F	F
V	V	F	F	F
F	V	V	F	F
V	F	F	F	V
F	V	V	V	F
V	V	F	V	F
F	F	F	V	V

6. Construa uma árvore de decisão, usando o algoritmo ID3, calculando a ganho de informação para cada nó. Inclua todos os passos do cálculo na resposta.
7. Considerando os atributos A1 a A4, aplique NaiveBayes como um algoritmo de aprendizado probabilístico e crie uma tabela com frequências e probabilidades para a coleção. Use a técnica de suavização de Laplace (ou seja, some 1 a todas as frequências) para evitar probabilidades 0.
8. Como o kNN classificaria o caso de teste $t1 = \{A1 = V, A2 = V, A3 = V, A4 = V\}$ considerando os atributos A1 a A4 usando 5 vizinhos ($k = 5$)? Assuma que a distância entre atributos simbólicos é 0 se eles têm os mesmos valores e 1, caso contrário. Use uma métrica de distância Euclidiana e calcule a classe sem ponderação (votação simples).