data

Aula 3

Nicolas de Sousa Maia
/NicolasSMaia

Presença

- Linktree: Presente na bio do nosso instagram
- Presença ficará disponível até 1 hora antes da próxima aula
- É necessário 70% de presença para obter o certificado

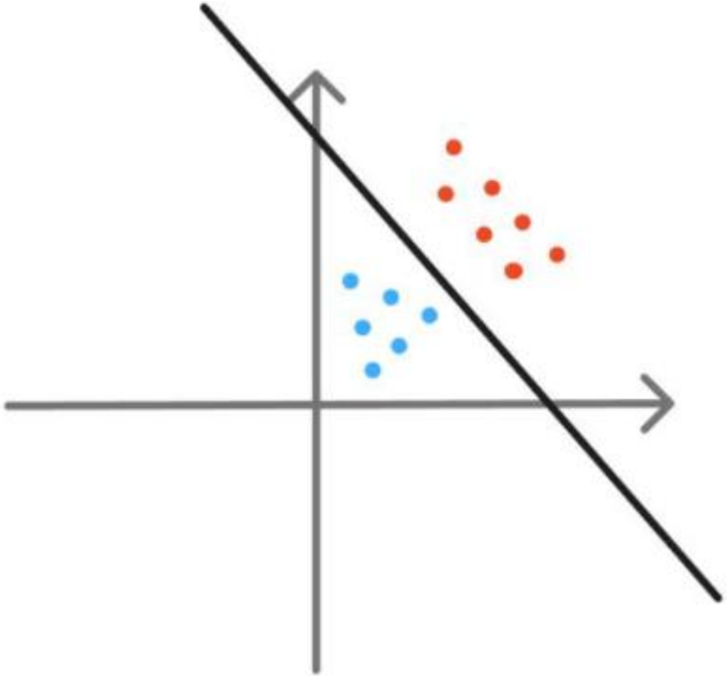
Recado

- Não teremos aula nos dias: 18/04, 02/05, devido aos feriados

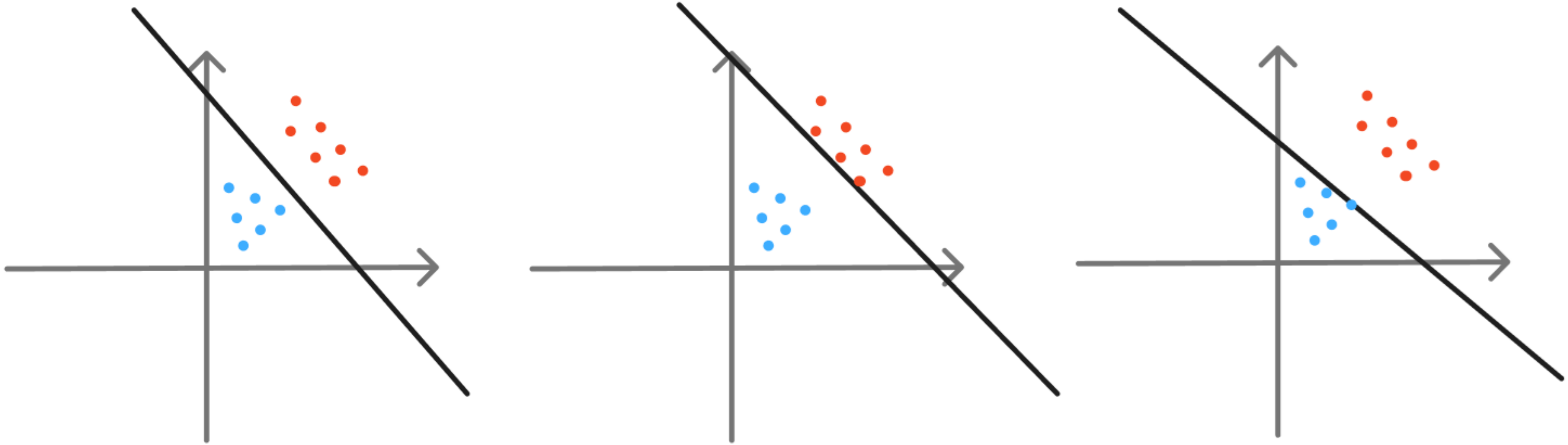
Presença e Github

Revisão

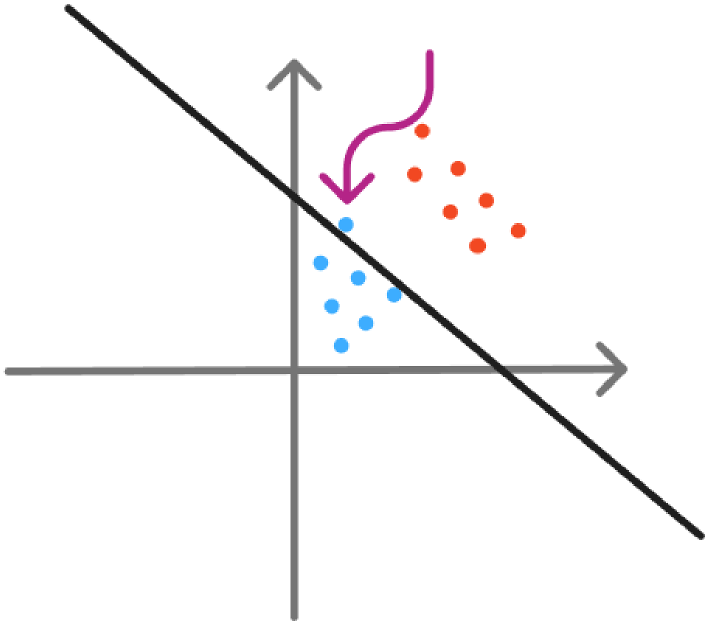
Classificação



Margem



Margem

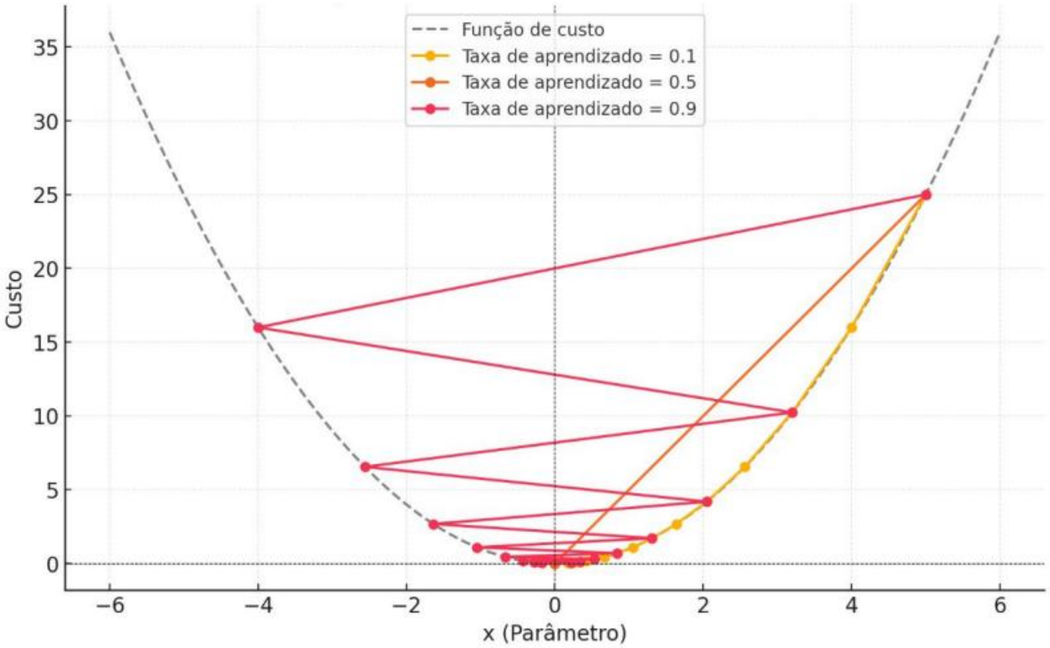


Função Objetivo

$$J(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} Loss(y_i(\vec{\theta} \cdot \vec{x} + \theta_0)) + \frac{\lambda}{2}\|\theta\|^2$$

∇

Gradiente



Função de custo
Taxa de aprendizado = 0.1
Taxa de aprendizado = 0.5
Taxa de aprendizado = 0.9
Custo
x (Parâmetro)
35
30
25
20
15
10
5
0
−6
−4
−2
0
2
4
6

Regressão Linear

Regressão x Classificação

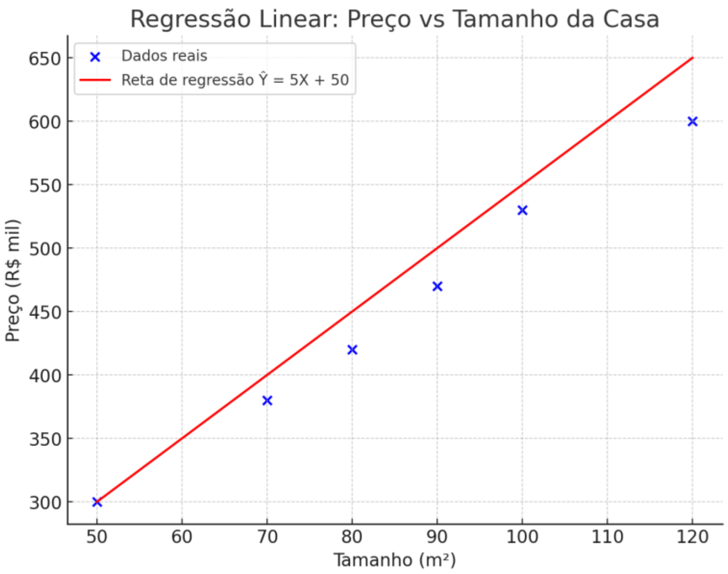
- Similaridade na forma
- Papel do hiperplano muda

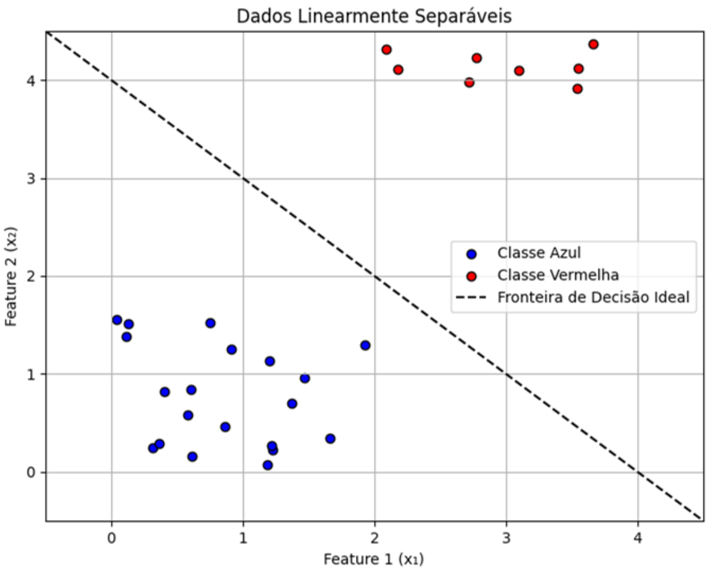
Objetivo da Regressão

- O objetivo está em prever um valor de saída para um novo valor de entrada, que pode ser 1 ou mais valores

Regressão x Classificação

- Colocando lado a lado, a diferença está no papel do hiperplano:





Regressão

Tabela 1: Dados de Tamanho e Preço de Imóveis

Casa	**Tamanho (m^2)** (X)	**Preço (R$ mil)** (Y)
1	50	300
2	70	380
3	80	420
4	90	470
5	100	530
6	120	600

Regressão

- O modelo da regressão linear segue:

$$Y = aX + b$$

- Encontramos os valores ótimos para o coeficiente angular e o coeficiente linear (chamado de intercepto)

Regressão

- Para o exemplo dado, temos:

$$\hat{Y} = 5X + 50$$

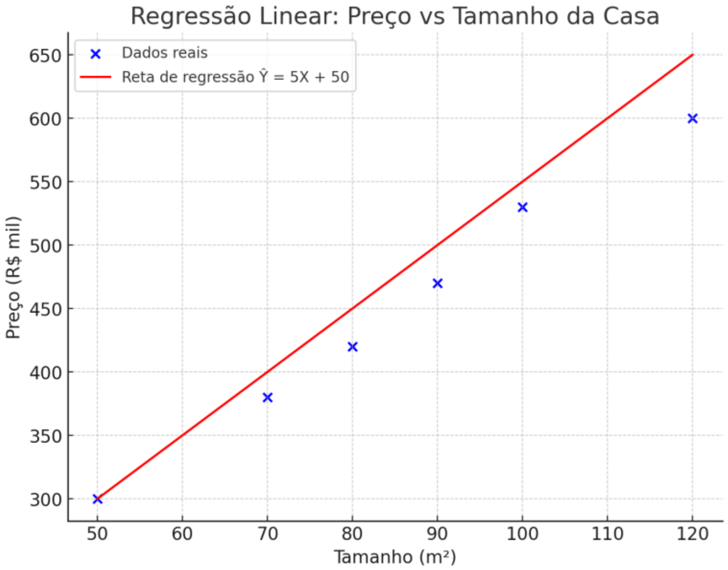
- Portanto, para um novo X, podemos prever com:

$$\hat{Y} = 5 \times 110 + 50 = 600 \text{ mil reais}$$

- A forma com que encontramos os parâmetros será explicada na sequência

Regressão

- Para visualização, temos o seguinte gráfico da regressão linear encontrada:



Mean Squared Error

Lousa

Mean Squared Error (MSE)

- Uma das principais Loss

$$L(f(x_i,\theta),y_i) = \frac{(y_i - f(x_i))^2}{2} = \frac{(y_i - \theta x_i)^2}{2}$$

- Intensifica a Perda, tanto para baixo quanto para cima

Detalhe!

$$\theta = (\theta_{antigo}, \theta_0)$$

Detalhe!

$$\theta = (\theta_{antigo}, \theta_0)$$

$$x = (x_{antigo}, 1)$$

Detalhe!

$$x \cdot \theta = x_{antigo} \cdot \theta_{antigo} + 1 * \theta_0$$

Empirical Risk

Empirical Risk

Função de Perda da Última Aula

Quantificava o erro baseado em um único dado

Empirical Risk

Função de Perda da Última Aula

Quantificava o erro baseado em um único dado

X

Empirical Risk

Quantifica o erro baseado em todos os dados de treinamento

Empirical Risk

- Ideia: Média Aritméticas das Perdas

$$R_n(f(x_i)) = \frac{1}{n}\sum_{i=1}^{n} L(f(x_i), y_i))$$

Mean Squared Error (MSE)

- Empirical Risk com MSE

$$R_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{(y^t - \theta x^t)^2}{2}$$

Regressão e Gradiente

Relembrando o Algoritmo do SGD

$$
\begin{aligned}
&1. \quad \theta = \text{valor aleatório}; \theta_0 = \text{valor aleatório} \\
&2. \quad \text{for } t \text{ in range}(T) \\
&3. \qquad \text{sortear i em: } \{x_0, x_1, \ldots, x_n\} \\
&4. \qquad\quad \theta = \theta - \eta \nabla_\theta [J]
\end{aligned}
$$

Gradiente do MSE

- Utilizando o MSE como Loss, podemos calcular seu Gradiente

$$\nabla L(f(x_i, y_i)) = \frac{\partial L(f(x_i, y_i))}{\partial \theta} = \frac{\frac{\partial (y_i - \theta x_i)^2}{2}}{\partial \theta} = -x_i(y_i - \theta x_i)$$

- Caminharemos na Direção Contrário ao Gradiente: $+x_i(y_i - \theta x_i)$

Gradiente do MSE

- Algoritmo de Atualização de θ

1. Inicializar $\theta = 0$
2. Tomar randomicamente um $t = 1, ..., n$
3. $\theta = \theta + \eta x^t (y^t - \theta x^t)$

Prática 1

Feature Transformation

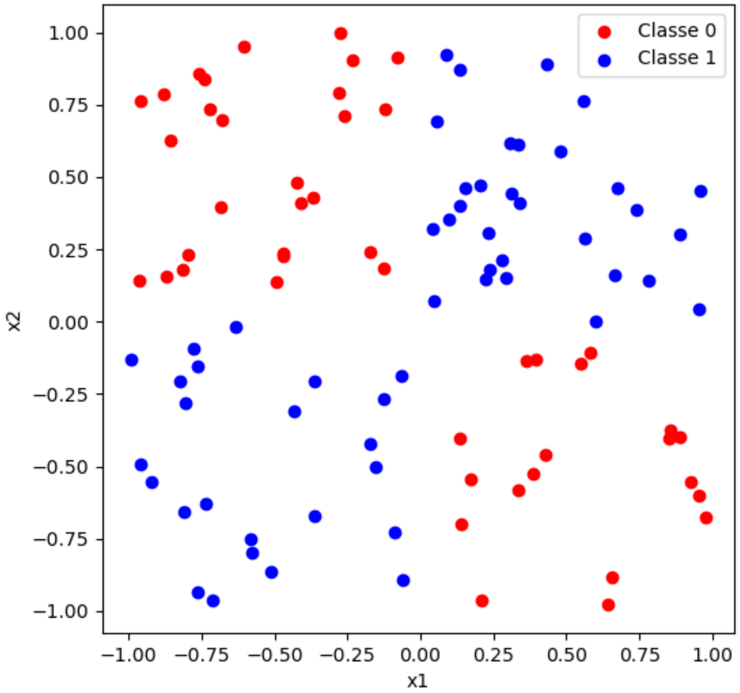
Definição

- Transformação matemática do vetor de features
- Tem como intuito mudar o espaço em que estamos trabalhando
- Pode ser de vários tipos

$$(x_1, x_2) \xrightarrow{f} (f(x_1), f(x_2))$$

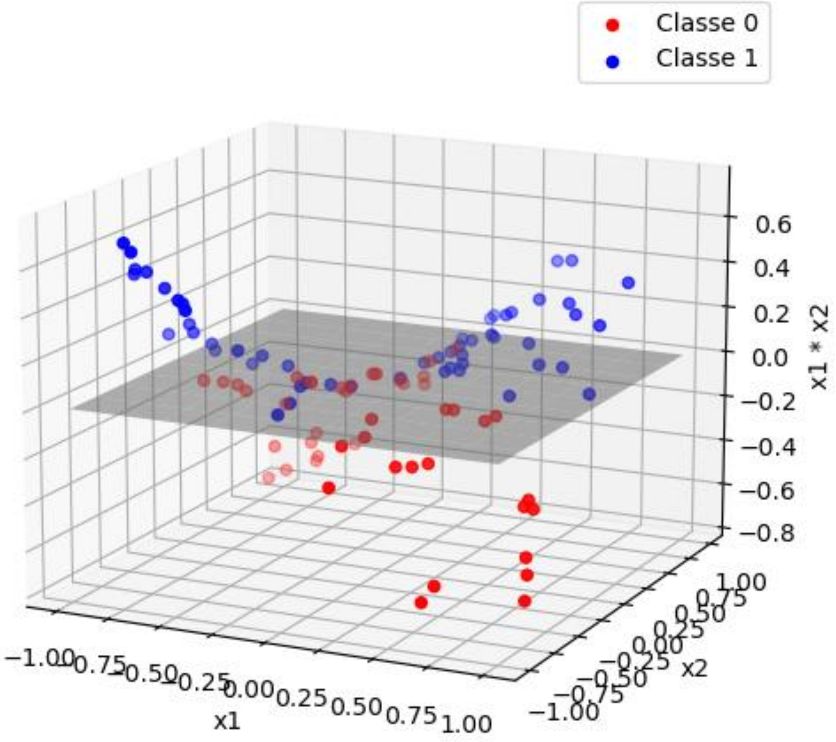
$$(x_1, x_2) \rightarrow (x_1, x_2, x_1^2, x_2^2)$$

Exemplo



Classe 0
Classe 1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
x2
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
x1

$$(x_1, x_2) \to (x_1, x_2, x_1 * x_2)$$



Transformação Polinomial

Transformação Polinomial

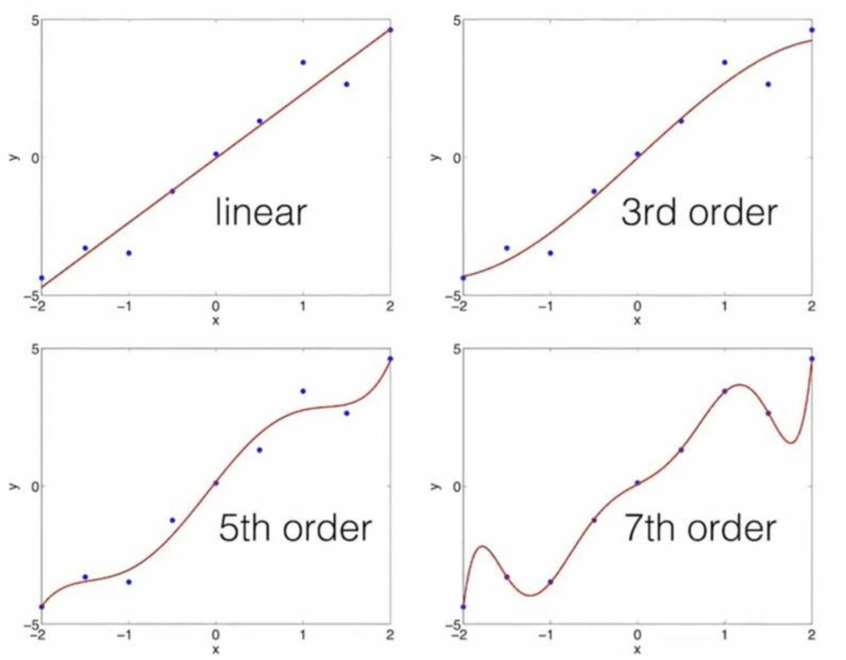
- É a listagem de todos os monômios, formados pelos elementos do nosso vetor de features, que os graus são menores ou iguais ao grau da expansão que queremos

Transformação Polinomial

- Primeira Ordem (Linear)

$$\phi(x) = [1, x]$$

- 1: necessário para θ_0 dentro de θ

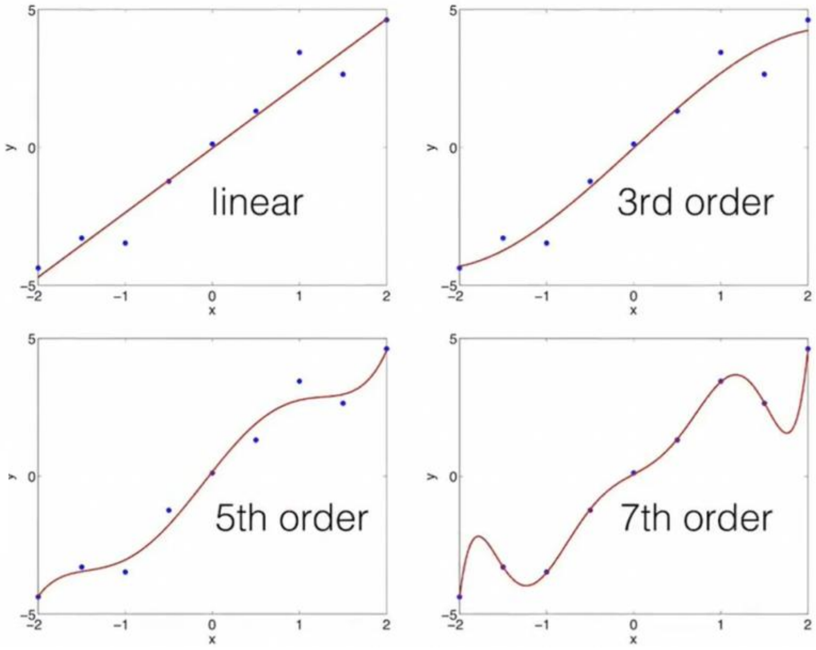


Transformação Polinomial

- Segunda Ordem (Quadrática)

$$\phi(x) = [1, x, x^2]$$

$$\phi_2(x) = [1, x_1, x_2, x_1^2, x_2^2, x_1 x_2]$$

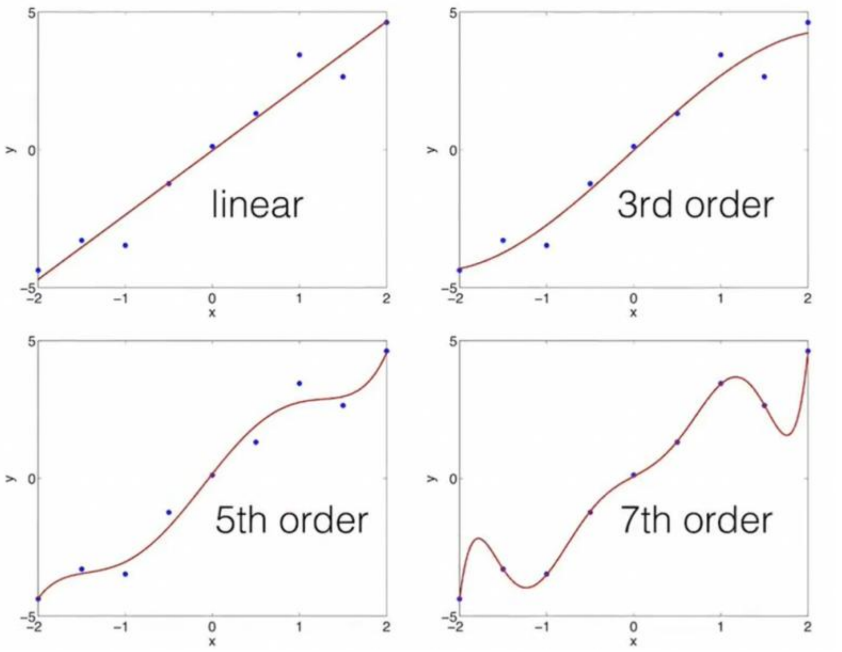


Transformação Polinomial

- Terceira Ordem (Cúbica)

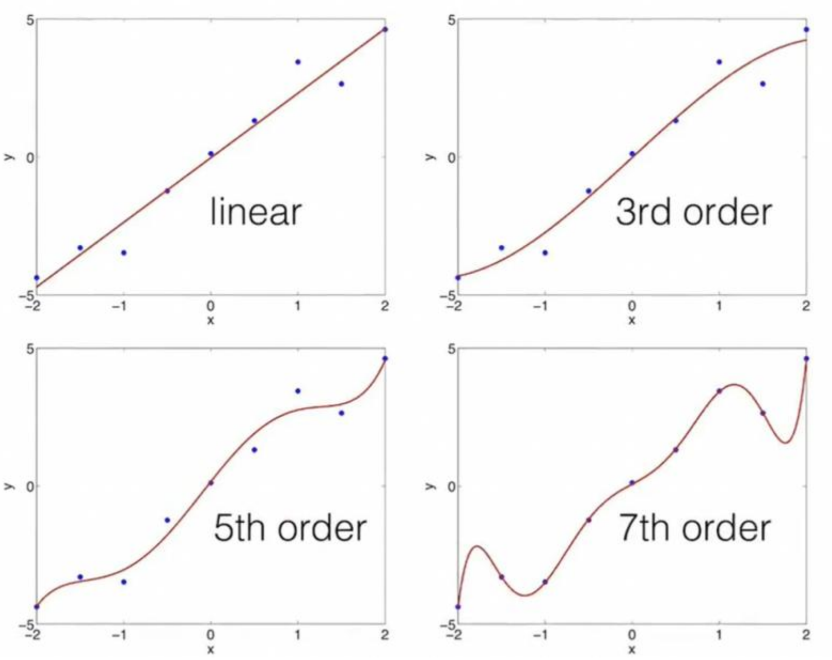
$$\phi(x) = [1, x, x^2, x^3]$$

$$\phi_3(x) = [1, x_1, x_2, x_1^2, x_2^2, x_1 x_2, x_1^3, x_2^3, x_1^2 x_2, x_1 x_2^2]$$



Alta Dimensionalidade

- Alta dimensionalidade pode causar overfitting
- Muito custoso para calcular

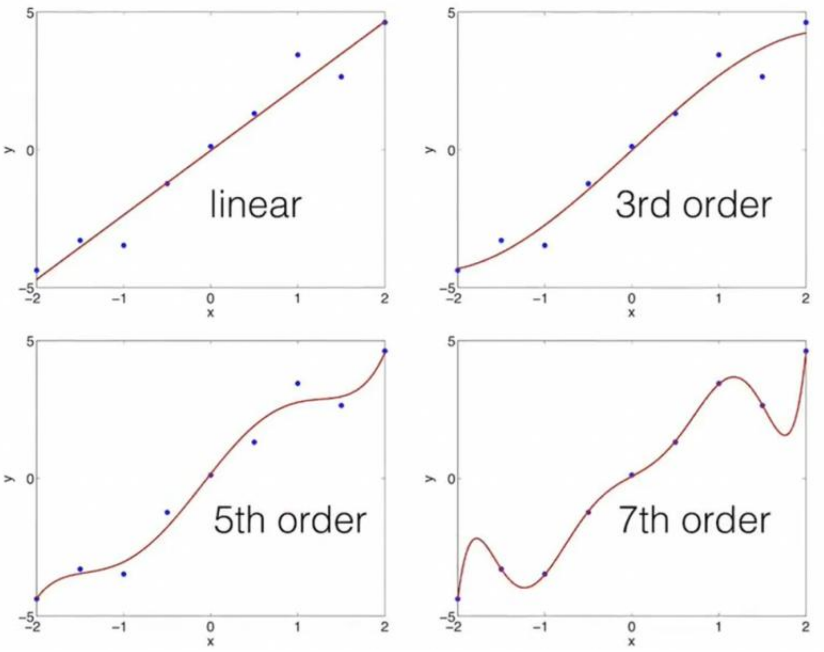


- Funções Kernel (próxima aula)

Regressão Polinomial

Regressão Polinomial

- Similar a regressão linear
- Vetor de características -> ϕ
- Vetor de parâmetros terá a mesma dimensionalidade de ϕ



Regressão Polinomial

1. Inicializar $\theta = 0$
2. Tomar randomicamente um $t = 1, ..., n$
3. $\theta = \theta + \eta\phi^t(y^t - \theta\phi^t)$

Prática 2

data@icmc.usp.br

@data.icmc

/c/DataICMC

/icmc-data

data.icmc.usp.br

obrigado por sua presença!