

# Anotações do Curso Statistical Learning - EDX - Capítulo 02

Rafael Barbosa da Silva

29 de maio de 2020

## 1 Introduction to Regression Models

### 1.1 Exemplo

- TV, Radio e Newspaper são variáveis que estão relacionadas com a Sales;
- Podemos prever a variável Sales usando a TV, Radio e Newspaper ?
- $\text{Sales} = \text{função}(\text{TV}, \text{Radio}, \text{Newspaper})$
- Sales: Variável Y
- TV, Radio e Newspaper: Variáveis X
- Matriz de features onde:  $X_1, X_2, X_3$  são as features

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

- Onde as dimensões da matriz são n (número de features), nesse caso  $n = 3$ , linhas e 1 coluna
- Podemos escrever nosso modelo como:  $Y = f(X) + E$ , onde E é o erro aleatório;
- Com a  $f(X)$  podemos:
  - fazer previsões para novos pontos de  $X = x$
  - Verificar quais das 3 variáveis são importantes para explicar a variável Y (Sales), quais seus impactos nesta.

## 1.2 Qual $f(x)$ é a ideal?

- Dado que temos 2 variáveis  $X$  e  $Y$ , qual valor seria o ideal de  $Y$ , para  $X = 4$ ;

$$f(4) = E(Y|X = 4)$$

- Isto é, a média dos valores que são igual a 4 nos dados que já temos
- Podemos ter muitos dados na qual  $X = 4$  corresponde a um valor de  $Y$ , nesse caso, estamos escolhendo a média destes valores e estamos lhe considerando um bom valor;

## 1.3 A função de regressão

- $f(x) = f(x_1, x_2, x_3) = E(Y|X_1 = x_1, X_2 = x_2, X_3 = x_3)$
- A melhor de  $f(x)$  é a que minimiza a função  $E[(Y - g(X))^2]$  para todos os pontos de  $X = x$ .
- O erro é uma expressão que pode ser definida por:  $E = Y - f(x)$ , é irreduzível;
- Se desenvolvermos a expressão acima, podemos escrever como:

$$E[(Y - g(X))^2] = [f(x) - \hat{f}(x)]^2 + Var(E)$$

## 1.4 Como estimar $f$

- na prática, temos poucos ou nenhum dado com  $X = 4$ , exatamente;
- Logo, não podemos computar  $E(Y|X = 4)$ , já que seria 0;
- Podemos fazer então:

$$\hat{f}(x) = Ave(Y|X \in N(x))$$

onde  $N(x)$  são os dados vizinhos a  $x$ ;

## 1.5 Dimensionality and Structured Models