

$\mathbb{R}^1 \rightarrow$ conj. pontos | $\mathbb{R}^2 \rightarrow$ linha | $\mathbb{R}^3 \rightarrow$ superfície | $\mathbb{R}^{n>3} \rightarrow$ hiper superfície

Modelo: coleção de funções preditivas

↳ **Treinar um modelo:** descobrir qual função dentro do modelo que tem melhor desempenho preditivo nos dados de **treinamento**

↳ **Algoritmo de treinamento:** um jeito sistemático de **treinar um modelo**

↳ **trade off viés/variação:** **underfitting** (underfitting/overfitting) modelo ruim tanto no **treino** quanto no **teste**

viés (bias): modelo é "temeroso", faz engarrafos que se sobrepõe aos dados

| | | | |
|---|--|---|---|
| underfitting ← \ominus flexível premissas fortes que "engarrafam" o modelo <div>desempenho tst ruim</div> | adequado <div>desempenho tst bom</div> | overfitting → \oplus flexível premissas tão fracas que "seguem as ridículas" do modelo <div>desempenho tst ruim</div> | overfitting: modelo faz flexível que se ajusta aos dados, desempenha de muito bem no treino e provavelmente mal no teste variação (variance): modelo é "influenciável" pela variação dos dados, resultado em um modelo MUITO bom (pouco provável) ou MUITO ruim (bem provável) |
|---|--|---|---|

Features ~ var. preditora

↳ atributos de análise de um modelo

Variáveis

↳ **Contínuas:** n° real (consegue fazer cálculos)

↳ **Discretas**

↳ **ordinais:** existe noção de **ordem**

↳ **nominais (categóricas):** **não** existe noção de **ordem**

Feature v.s Representação (forma da feature que o modelo lê)

| Feature | Representação |
|------------|--|
| contínua | o próprio valor |
| categórica | variáveis "dummy" |
| ordinal | converte p/ contínua ou (one hot encoding) categórica |