

# Aprendizado supervisionado

Prof. Hugo de Paula



# Separação das bases de dados



# Bases de treinamento e de testes

- Devem ser semelhantes (estatisticamente, cobertura do espaço de solução, etc.)
- Se base de dados é grande, pode-se partir a base (*percentage Split*).
- Caso contrário, usar validação cruzada (*cross validation*).



# Bases de treinamento e de testes

## ■ *Percentage Split / Hold out*

- Particiona a base por amostragem.
- Tipos de amostragem interferem no resultado.

## ■ *Cross Validation*

- Particiona em K-partes.
- Por exemplo: Separa a base em 10 partes. Em cada rodada usa 9 blocos para treinamento e 1 bloco para teste.





