



ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

## PQI 3403 Análise de Processos da Indústria Química

Ardson dos Santos Vianna Júnior - ASVJ  
e-mail: [ardson@usp.br](mailto:ardson@usp.br)





ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

# Aula 9\_3

## Machine learning

PQI 3403 Análise de Processos da Indústria Química



# Sumário

---

Contextualização

---

Tipos de Aprendizagem

---

Quem? Atributos

---

Como? Treinamento

---

Regressão

---

Conclusão



# 1. paradigma básico

- Observar um conjunto de exemplos: treinamento de dados
- Inferir algo sobre o processo que gerou os dados
- Usar a inferência para fazer previsões sobre dados que não foram avaliados: teste de dados



## 2 Quem? Representação

- Engenharia de atributos

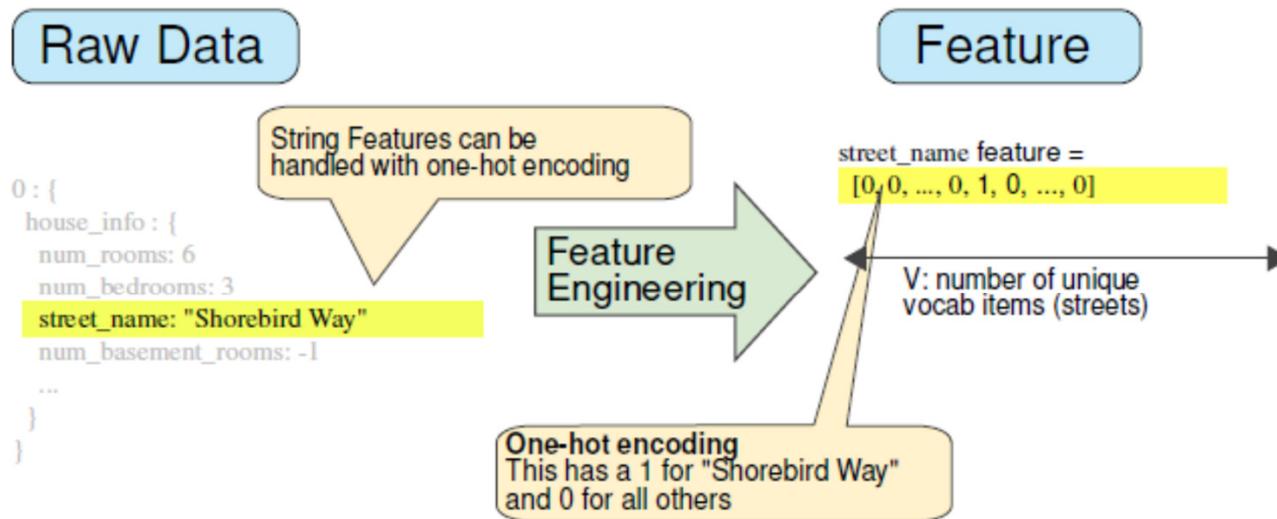


- Preferência por números



## 2 Representação

- **Atributos categóricos:** ex. nomes de rua' {‘Luciano Gualberto’, ‘Lineu Prestes’, ‘Matão’} converter a números 1,2,3



## 2 Representação – boas práticas

- Evitar valores raramente usados
- Preferir significados claros e óbvios
- Não misturar valores “mágicos” com valores reais – difícil interpretação



## 2. Representação

- **Escalonamento** (scaling): converter números de sua faixa natural a faixas padrões, 0 a 1, -1 a +1
- Tratar *outliers* extremos.
- **Binning**: empacotamento - faixas
- **Scrubbing**: depuração
- **Representação esparsa**: se muitos zeros, só guardar os valores diferentes de zero.



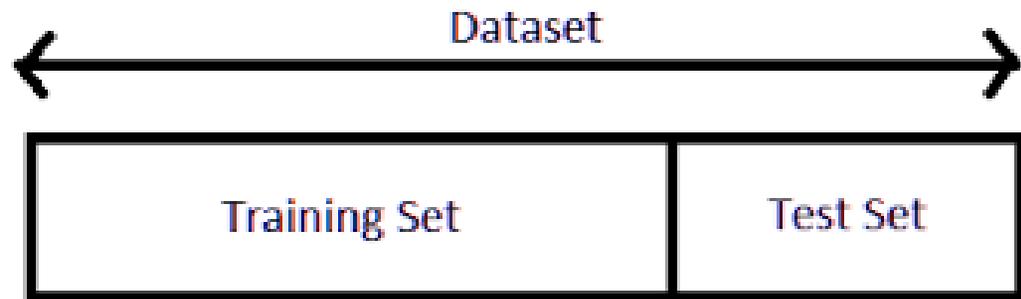
## 3. Como? Treinamento

- Primeiro:
  - um padrão é apresentado à camada de entrada da rede. A atividade resultante flui através da rede, camada por camada, até que a resposta seja produzida pela camada de saída.
- No segundo passo:
  - A saída obtida é comparada à saída desejada para esse padrão particular. Se esta não estiver correta, o erro é calculado.
  - O erro é propagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retropropagado.



### 3. Treinamento e teste – dividindo os dados

- Dados para treinamento
- Dados para teste



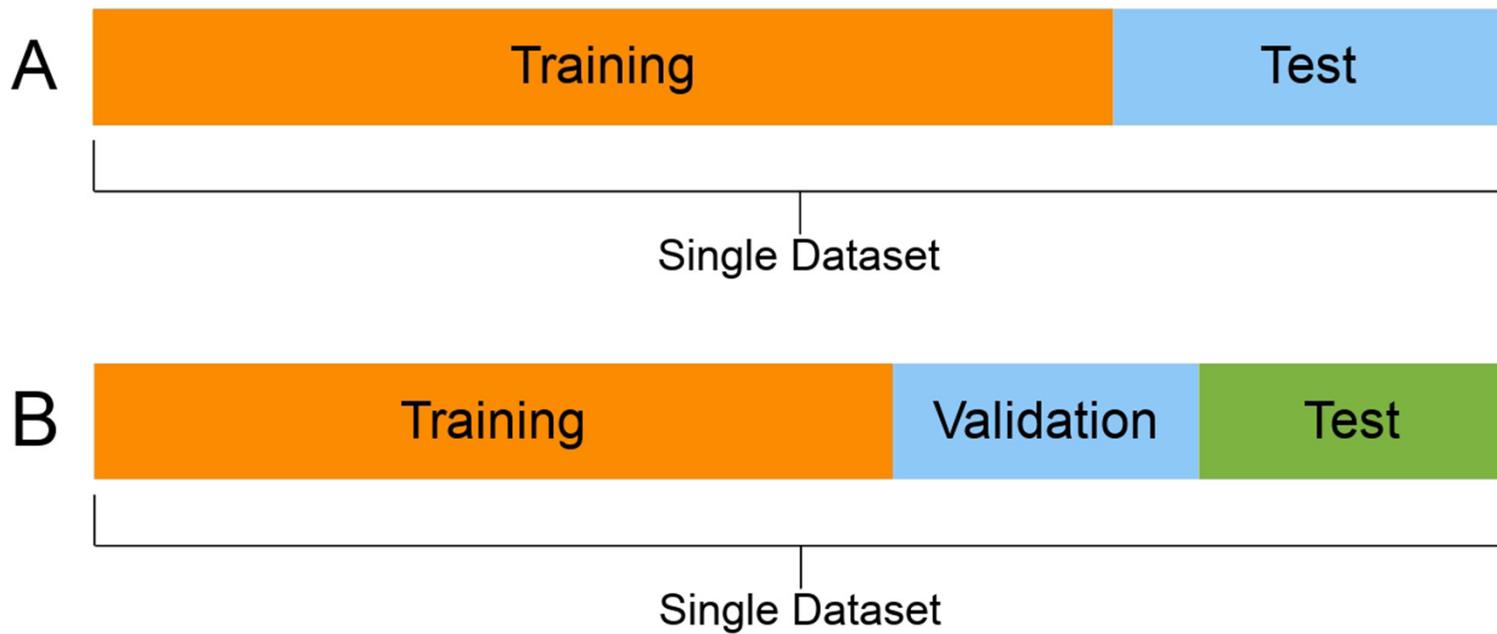
### 3. Treinamento e teste – dividindo os dados

- Dados para treinamento:  
grande o suficiente para ter significado estatístico
  
- Dados para teste  
Não incluir nos dados para treinamento



### 3. Melhorando a partição

- Conjunto de dados para validação





# Conclusão

- Engenharia de atributos
- Partições do conjunto de dados



# Bibliografia

- <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-867-machine-learning-fall-2006/index.htm>
- Peter Norvig, Stuart Russell - Inteligência Artificial, tradução da 3a ed., Elsevier, 2013.
- <https://developers.google.com/machine-learning/crash-course/ml-intro>
- <https://www.youtube.com/watch?v=h0e2HAPTGF4>