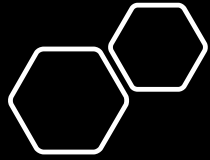


Ajuste de Hiper Parâmetros

- Um classificador pode ter diversos hiper parâmetros
- A configuração dos hiper parâmetros as vezes não é intuitiva e nem “científica”
- A combinação de muitos hiper parâmetros pode ser gigantesca
- Inviável testar manualmente todas as combinações

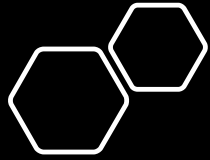
Ajuste de Hiper Parâmetros

- Spark facilita o teste de diferentes hiper parâmetros
- Você pode testar diferentes configurações para todo o Pipeline
- ParamGridBuilder: você configura quais parâmetros e quais domínios quer testar
- Utiliza Cross-Validation ou Train-Validation Split



Cross- Validation

- Divide os dados de treino em partições (folds) que são usadas para treino ($2/3$ dos dados) e teste ($1/3$ dos dados)
- 3 Partições = 3 conjuntos de treino e 3 de teste
- As partições são treinadas e testadas nos domínios do grid de hiper parâmetros fornecido
- Quando encontrar o melhor conjunto de hiper parâmetros, ele novamente cria o modelo usando todos os dados de treino



Train- Validation Split

- Testa cada configuração do grid de hiper parâmetros uma única vez
- `trainingratio` define a proporção usada para treino
- Quando encontrar o melhor conjunto de hiper parâmetros, ele novamente cria o modelo usando todos os dados de treino

	Custo Computacional	Resultados
Cross-Validation	↑	↑
Train-Validation Split	↓	↓

Cross-Validation VS Train-Validation Split

