

## Comparação entre os classificadores

### OneClassSVM:

score nos genes de treinamento (genes reais + sinéticos): 0.9681528662420382

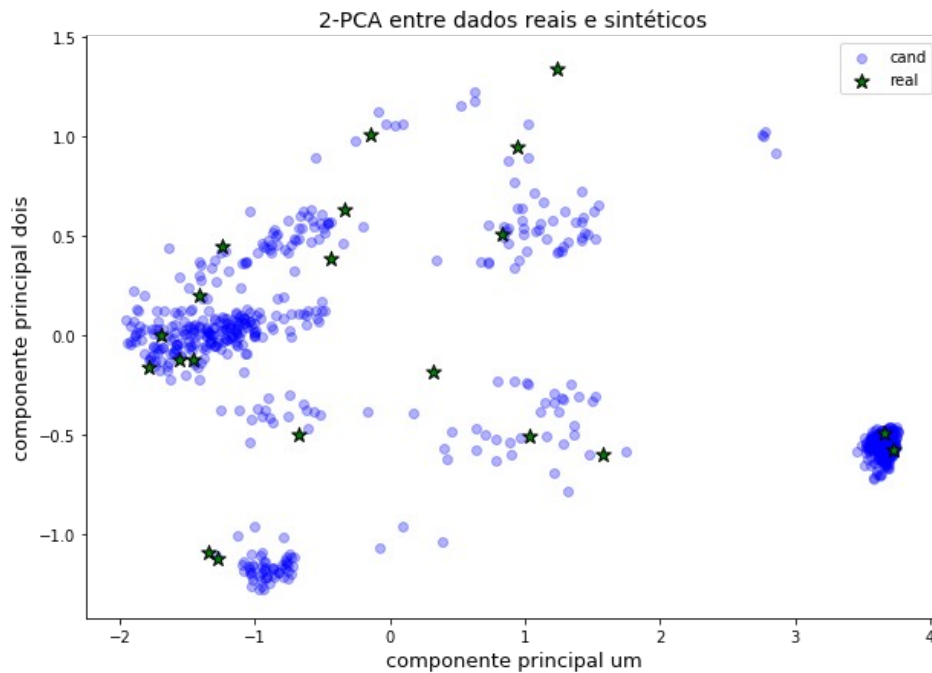
score nos genes de referência de treinamento(só reais): 0.9285714285714286

score nos genes de referência no utilizados pro treinamento: 0.8333333333333334

score em todos os genes de referência (train e test): 0.9

Redução do 82.94887039239 % no processo de escolha de candidatos, ou seja, ele acha que o 17% dos genes não classificados na literatura, são possíveis candidatos a genes de referência.

PCA 2-componentes com os genes que o classificador (svm) acha que são candidatos (azules)



### Isolation-Forest:

score nos genes de treinamento (genes reais + sinéticos): 1.0

score nos genes de referência de treinamento(só reais): 1.0

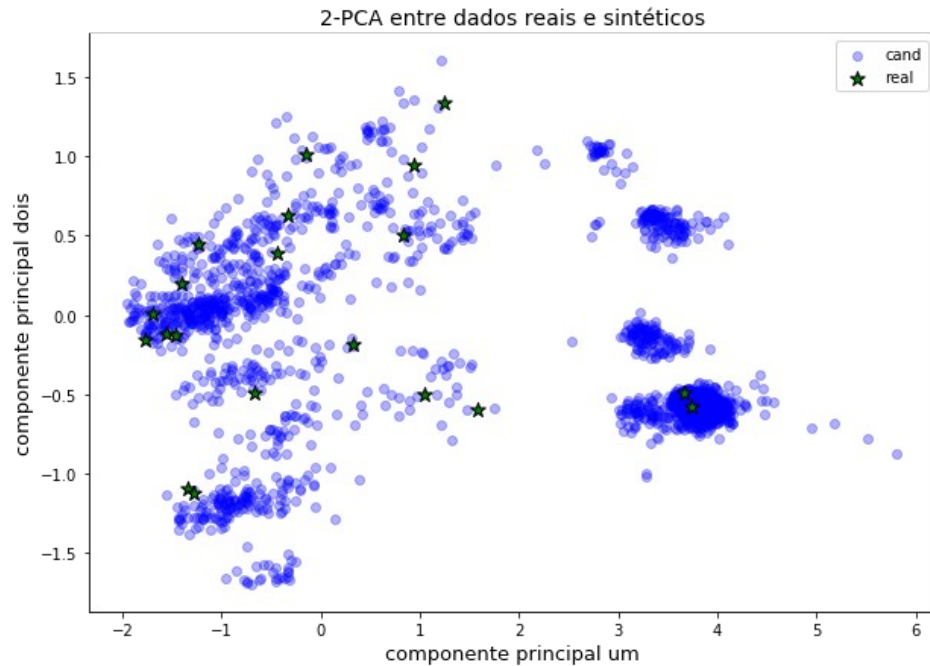
score nos genes de referência no utilizados pro treinamento: 0.8333333333333334

score em todos os genes de referência (train e test): 0.95

Redução do 47.990487514863254 % no processo de escolha de candidatos, ou seja, ele acha que o 52% dos genes não classificados na literatura, são possíveis candidatos a genes de referência.

Dica: Como ele cria umas fronteiras baseadas no intervalo das características de treinamento, e os dados de testes são muito semelhantes, então a fronteira fica fraca.

PCA 2-componentes com os genes que o classificador (RandomForest) acha que são candidatos (azules)



### Conclusões:

- Eu acredito que é melhor usar o **SVM** já que a fronteira que a **RandomForest** faz é fraca, permite muitos genes devido a que os valores dos genes de referência e os genes não classificados é semelhante.