

BRUNO COUTO MARINO

Ajustar e Avaliar o Modelo

Ajustar e avaliar propor o modelo (usar Grid Search, Randomized Search ou métodos de Ensemble para ajustar o modelo; avaliar o modelo usando o conjunto de testes)

Estudos sobre Inteligência Artificial



Hiperparâmetros

Introdução

- Parâmetros de modelos que devem ser definidos antes de treiná-lo.
- A escolha deles interfere na acurácia final

Hiperparâmetros

Introdução

- Parâmetros de modelos que devem ser definidos antes de treiná-lo.
- A escolha deles interfere na acurácia final
- Exemplos de hiperparâmetros:
 - Random Forest - número de árvores
 - Redes Neurais - taxa de aprendizado, número de camadas, quantidade de nó em cada camada, etc.
 - SVM e SVR - c , gamma e epsilon.

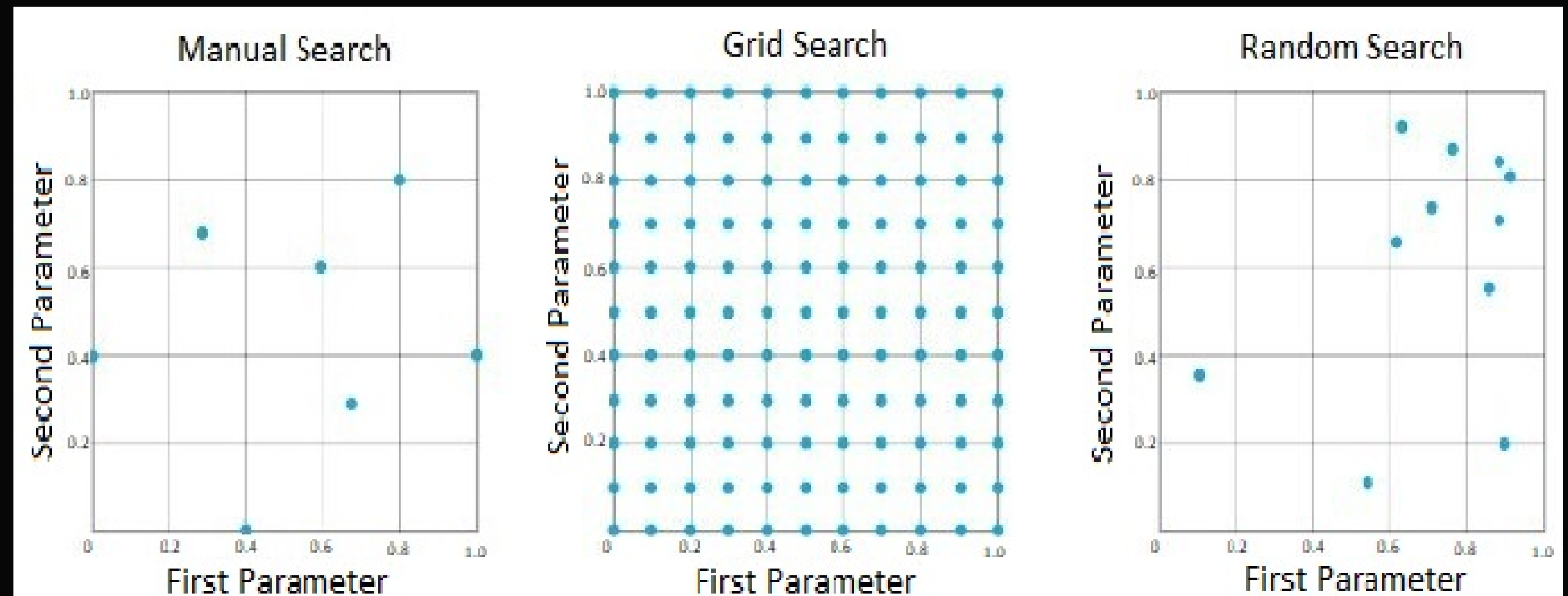
Como escolher os valores dos Parâmetros?

Grid Search

Testar todas as testar combinações possíveis dos hiperparâmetros, exaustivamente

Randomized Search

Testa aleatoriamente as combinações dos hiperparâmetros.



Método Ensemble

AGRUPAMENTO

Calcula a previsão final baseada em previsões de diversos modelos.

Treina modelos mais simples para uma mesma tarefa e produz um modelo mais complexo

Random Forest

Método Ensemble

AGRUPAMENTO

Calcula a previsão final baseada em previsões de diversos modelos.

Treina modelos mais simples para uma mesma tarefa e produz um modelo mais complexo

Random Forest

PROBLEMAS EXISTENTES

Necessidade de modificar o Dataset

Poucos dados de treino

Bootstrapping - retirar amostras aleatórias com reposição

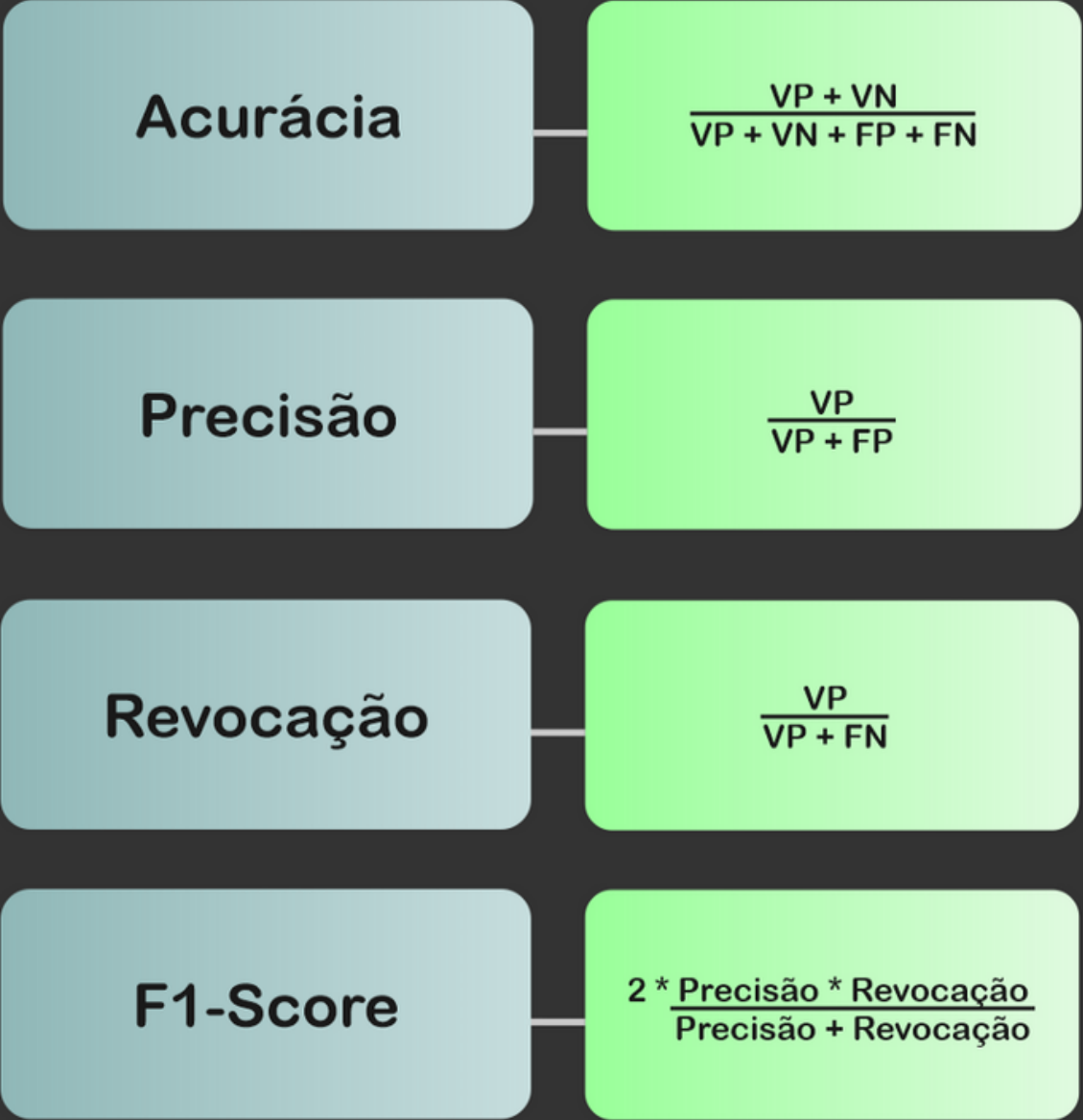
Bootstrap aggregating

Avaliar o Modelo

MATRIZ DE CONFUSÃO

Visualizar os resultados do modelo

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)



Acurácia

Performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente;

Recall/Sensibilidade

Das classificações "Positivo" como valor esperado (real), quantas estão corretas;

Precisão

Das classificações "Positivo" estimadas pelo modelo, quantas estão corretas;

Precisão

Média harmônica entre precisão e recall.

Liks possivelmente úteis:

<https://gusrabbit.com/code/grid-search/>

<https://www.sciencedirect.com/topics/mathematics/grid-search>

<https://medium.com/data-hackers/otimizando-os-hiperpar%C3%A2metros-621de5e9be37>

<https://medium.com/turing-talks/turing-talks-18-modelos-de-predi%C3%A7%C3%A3o-random-forest-cfc91cd8e524>

https://www.bioinfo.ufpr.br/moodle/pluginfile.php/725/mod_resource/content/0/05_matriz_de_confusao.pdf

<https://vitorborbarodrigues.medium.com/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-recall-quais-as-diferen%C3%A7as-c8f05e0a513c>

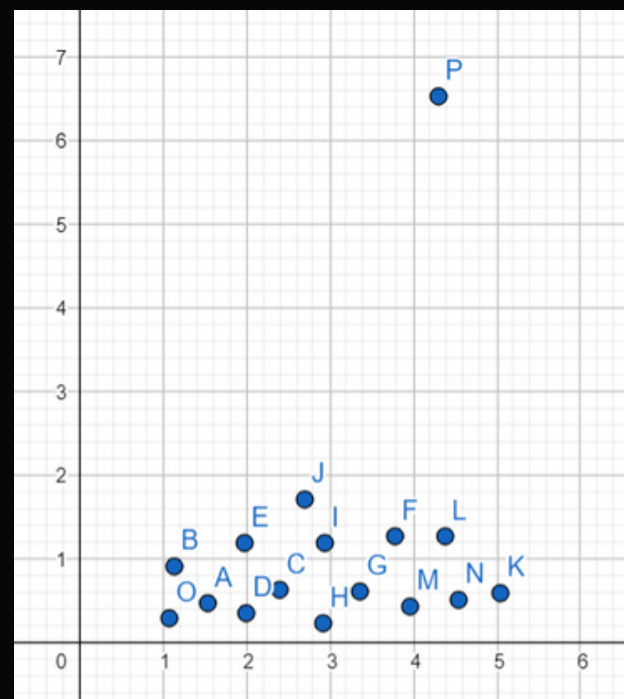
Método ense

Valores Ausentes

Não necessariamente é ruim

Podemos substituir com médias ou medianas.

Outliers



Pode ser justamente o que está sendo procurado.

Normalização dos Dados

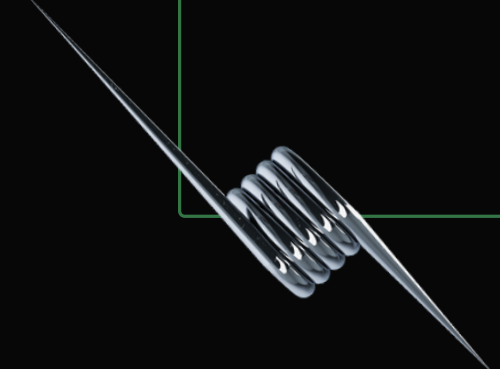
Ajustes na escala

Padronizar variáveis contínuas (média 0 desvio padrão 1)

One-hot encoding



Subajustamento (underfitting)

- O modelo não conseguiu aprender o suficiente.
 - Leva à um erro elevado tanto nos dados de treino quando nos dados de teste.
- 

Sobreajustamento (overfitting)

- O modelo aprende demais sobre os dados.
 - Ele é adequado apenas para os dados de treino, como se não fosse capaz de generalizar para outros dados nunca vistos antes
- 