

# WiFi Signal Strength

Os serviços de localização são, quase que exclusivamente, baseados na tecnologia GPS. Muitas estratégias foram adotadas para fornecer aos usuários serviços personalizados baseados em GPS. Entretanto, o WiFi poderia ser explorado para fornecer tais serviços, como foi proposto por Rohra et al.<sup>1</sup>, visto que o sinal de satélites (GPS) pode ser demasiadamente fraco dentro de prédios e outras construções. Em linhas gerais, o método proposto por Rohra et al. consiste em classificar em qual cômodo o receptor se encontra baseado na intensidade (dBm) de sinais de WiFi gerados por sete roteadores distintos. Esse desafio foi solucionado com o uso de redes neurais (uma sub classe de inteligência artificial), as quais foram capazes de aprender a interpretar os valores de intensidade dos sinais recebidos.

Um requisito quase que mandatório para o aprendizado de máquina é termos um vasto banco de dados. Portanto, para o trabalho de Rohra et al. foi necessário coletar uma quantidade considerável de amostras reais em diversos instantes de tempo e em pontos diferentes no espaço. Esses dados coletados estão disponíveis no arquivo data.txt, no qual as sete primeiras colunas (da esquerda para a direita) são os valores da intensidade de sinal recebido, em dBm, de cada um dos sete roteadores WiFi e a última coluna da direita representa o cômodo no qual foi realizada a medida.

A tabela a seguir ilustra as 5 primeiras linhas dos dados.

Router 1	Router 2	Router 3	Router 4	Router 5	Router 6	Router 7	Room
-64	-56	-61	-66	-71	-82	-81	1
-68	-57	-61	-65	-71	-85	-85	1
-63	-60	-60	-67	-76	-85	-84	1
-61	-60	-68	-62	-77	-90	-80	1
-63	-65	-60	-63	-77	-81	-87	1

Em todos os exercícios a seguir deve-se utilizar os dados do arquivo data.txt. Nos anexos encontram-se alguns comandos úteis para a realização dos exercícios e um breve tutorial de como importar os dados.

---

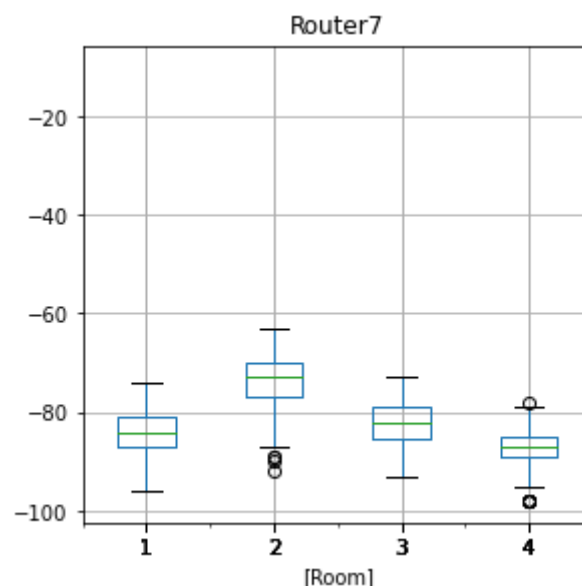
<sup>1</sup> Jayant G Rohra, Boominathan Perumal, Swathi Jamjala Narayanan, Priya Thakur, and Rajen B Bhatt, 'User Localization in an Indoor Environment Using Fuzzy Hybrid of Particle Swarm Optimization & Gravitational Search Algorithm with Neural Networks', in Proceedings of Sixth International Conference on Soft Computing for Problem Solving, 2017, pp. 286-295.

Exercício 1-) Os dados da tabela acima sugerem que o sinal gerado pelo roteador 2 é o mais intenso dentro da sala 1. Será mesmo isso verdade? Qual a forma mais simples de inferirmos qual é o sinal mais intenso dentro da sala 1? Apresente os cálculos necessários juntamente com a sua conclusão. Refaça o exercício para a salas 2, 3 e 4.

Exercício 2-) Agrupe os dados por salas (Room) e represente graficamente os *boxplot* para os sete roteadores. Observando os gráficos, podemos sugerir uma forte correlação entre os roteadores 6 e 7. Calcule a correlação par-a-par entre todos os roteadores usando todos os dados disponíveis. Realmente existe uma forte correlação entre os sinais gerados pelo roteador 6 e pelo roteador 7? Outro par de roteadores apresenta um maior coeficiente de correlação? Dica: A maioria dos programas já possui um comando para o cálculo dos coeficientes de correlação entre sinais. Para um conjunto de sinais, o resultado será a matriz de correlação.

Exercício 3-) Suponha que para um melhor desempenho da inteligência artificial seja necessário dispensarmos as amostras de um roteador. Admita, também, que o critério utilizado para selecionar o sinal do roteador que será dispensado seja: dispense as amostras que, em média, mais se desprendam/distanciem da sua média amostral. Assim sendo, determine qual roteador (e conseqüentemente todas as suas amostras) deve ser dispensado em cada sala. Mostre como você realizou os cálculos.

Exercício 4-) Gere sete gráficos do tipo *boxplot*, cada gráfico será constituído pelos dados de um roteador. Cada gráfico terá 4 'pontos'/símbolos, um para cada sala. Para que fique bem claro, apresentamos a seguir o resultado desejado para o Router7.



Admita que os dados que apresentem uma menor variabilidade entre as salas sejam menos relevantes ao algoritmo de detecção. Diante dessa suposição e com base nos gráficos que você acabou de gerar, qual(is) dos sinais/routers você julga desnecessário(s)?

Exercício 5-) Gere um histograma normalizado (frequência relativa) para cada um dos sete sinais amostrados. Para cada um dos histogramas, ajuste a média e o desvio padrão de uma distribuição Gaussiana (também conhecida como distribuição normal) de forma que ocorra o melhor encaixe visual dos dados com a distribuição. Nesse processo de ajuste será necessário alterar o número de bins (intervalos de classes ou barras) em cada um dos histogramas. Deve-se utilizar no mínimo 5 bins e no máximo 15 bins. Feito isso, qual a sua opinião sobre aproximarmos os dados por distribuições Gaussianas? Dica: Você pode implementar a função densidade de probabilidade teórica de uma Gaussiana, ou mesmo usar uma função pronta (*built-in function*) disponível na maioria dos programas.

Exercício 6-) Crie uma variável aleatória correspondente ao nível médio dos roteadores, ou seja, essa nova variável aleatória é dada por

$$M = \frac{R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7}{7}$$

Na qual  $R_x$ , com  $x \in \{1, 2, \dots, 7\}$ , representa o nível de sinal recebido do roteador  $x$ . Gere o histograma de  $M$  com 20 bins e ajuste a média e o desvio padrão de uma distribuição Gaussiana de forma que ocorra o melhor encaixe visual do histograma com a distribuição Gaussiana. Admita a função densidade de probabilidade (FDP) abaixo

$$f_X(x) = \rho \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu_1}{\sigma_1} \right)^2} + (1 - \rho) \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu_2}{\sigma_2} \right)^2}$$

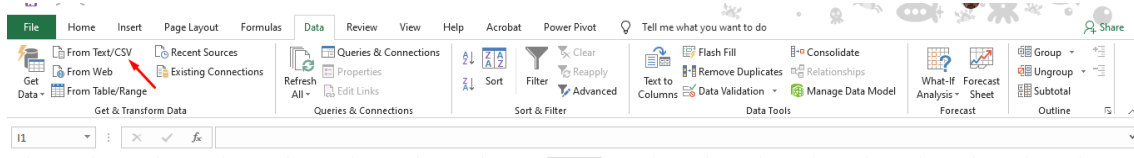
Podemos interpretá-la como a soma ponderada pelo fator  $\rho$  de duas distribuições Gaussianas distintas. Gere um novo histograma da variável aleatória  $M$  com 20 bins e ajuste os parâmetros da função  $f(x)$  para que ocorra o melhor encaixe visual com histograma. Qual dentre as duas funções, Gaussiana e  $f_X(x)$ , melhor se ajustou ao histograma?

Exercício 7-) Um breve resumo sobre o teorema central do limite pode ser encontrado em: [https://pt.wikipedia.org/wiki/Teorema\\_central\\_do\\_limite](https://pt.wikipedia.org/wiki/Teorema_central_do_limite). Após a leitura, discorra sobre o exercício 6. Qual deveria ser a melhor função, dentre as apresentadas no exercício 6, para aproximar a variável aleatória  $M$ ? Ela realmente foi melhor? Se não, qual poderia ser o motivo?

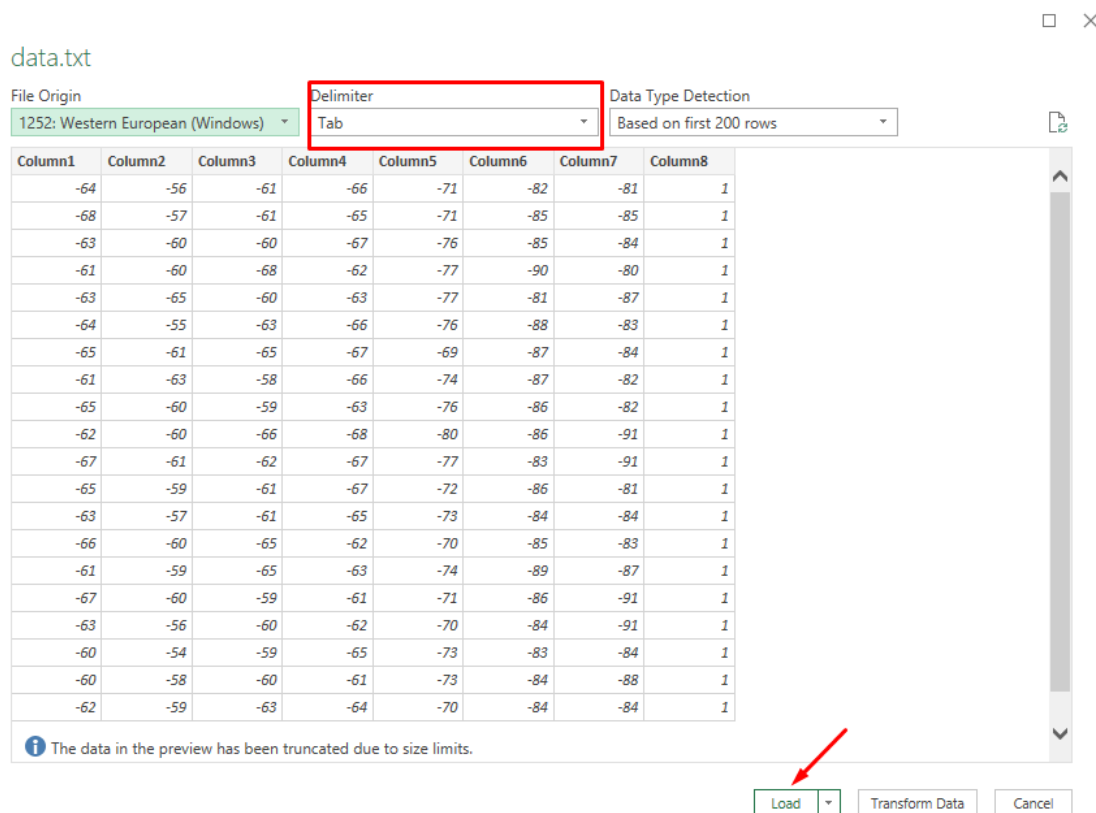
## Anexos

### Importando os dados no Excel:

Passo 1: Na aba Data selecione a opção From Text/CSV, conforme a figura abaixo.



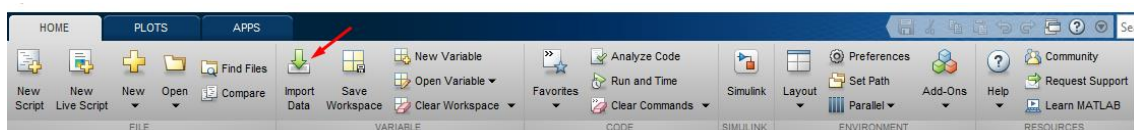
Passo 2: Após selecionar o caminho correto do arquivo data.txt, a seguinte aba será aberta.



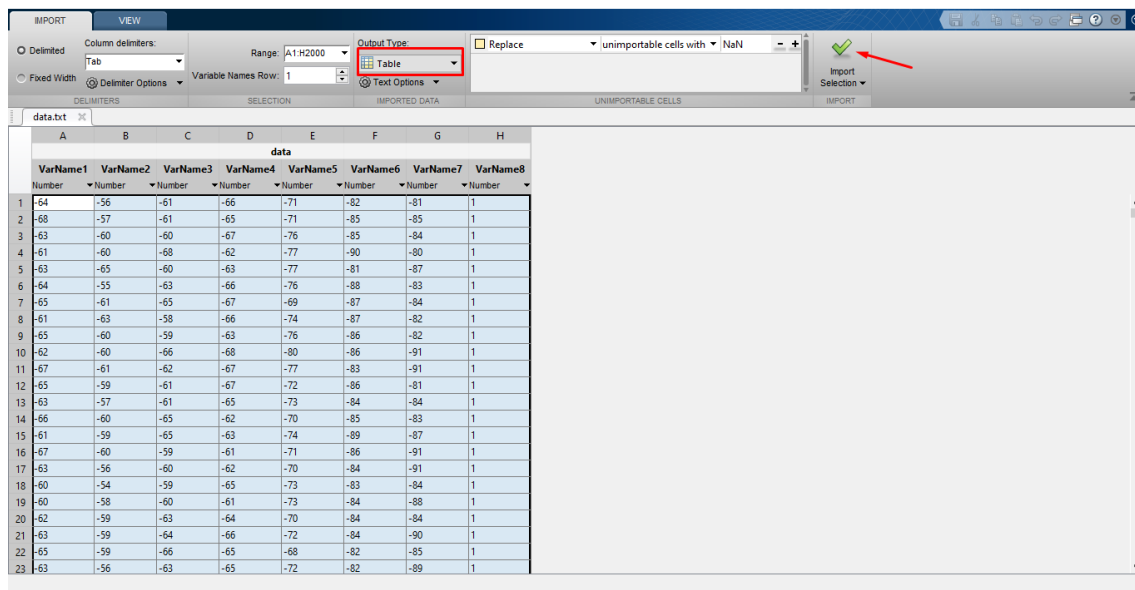
Confira o campo “Delimiter” e click em Load. Pronto, os dados foram carregados.

### Importando os dados no Matlab:

Passo 1: Na aba HOME do matlab selecione a opção “Import Data”, conforme a figura abaixo.



Passo 2: Após selecionar o caminho correto do arquivo data.txt, a seguinte aba será aberta.



Selecione o tipo de que você deseja (quadro vermelho), aconselhamos utilizar o tipo Table, e em seguida click em “Import Selection”. Pronto, os dados foram carregados em uma variável no workspace.

### Importando os dados no Jupyter (Python):

Aconselhamos usar o pacote pandas para manuseio dos dados. Utilize as seguintes linhas de código para importar os dados:

```
import numpy as np
dataframe = pd.read_fwf(path+'/data.txt',header=None,
names=["Router1", "Router2", "Router3", "Router4","Router5","Router6",
"Router7", "Room"])
```

Para visualizar os dados, por exemplo as 5 primeiras linhas:

```
dataframe.head(5)
```

Para selecionar dados no dataframe, por exemplo os dados da sala 2 ou sala 3:

```
a = dataframe.query('Room==2 | Room==3')
```

Mais informações podem ser encontradas em

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html>