

Para saber mais

Como podemos ver na configuração inicial da camada Dense na documentação do Keras,

```
keras.layers.Dense(units, activation=None, use_bias=True, kernel_initializer='glorot_uniform', bias_
```

o inicializador de kernel ou matriz de pesos é chamado de **glorot uniforme** e os vieses são inicializados com 0.

Glorot, pois é o sobrenome de Xavier Glorot que escreveu o artigo que define esse inicializador. Ele também é conhecido como o **inicializador uniforme de Xavier**.

E o que esse inicializador faz? Bom, ele pega nossos dados de treino e define possíveis valores para os pesos, como abaixo:

```
pesos = [1,2,3,4,5]
```

A cada propagação da rede esses valores alteram, sendo que a probabilidade de cada um deles acontecer permanece 1/5.

Fazer com que essa probabilidade permaneça igual para cada um dos números é algo chamado de **distribuição uniforme**.

Então, o algoritmo pega essas amostras de uma distribuição uniforme, define valores, mas quais valores? De 1 até 1.000? De 0 até 5? Qual o limite?

Para saber isso, o algoritmo resolve a seguinte conta:

```
sqrt(6 / (unidades_de_entrada_do_kernel + unidades_de_saida_do_kernel ))
```

Repare que ele pega tudo que entra para ser calculado, junta com tudo o que vai sair, que são nossas unidades de entrada e saída da camada, divide por 6 e pega a raiz quadrada desse número.

E porque isso é importante? Porque desse modo os pesos ficam dentro de uma faixa de valores, o que garante que eles fiquem “na medida”, nem muito pequenos a ponto de sumirem e nem muito grandes a ponto de explodirem.

Link do artigo (ou paper): <http://jmlr.org/proceedings/papers/v9/glorot10a/glorot10a.pdf>
(<http://jmlr.org/proceedings/papers/v9/glorot10a/glorot10a.pdf>)