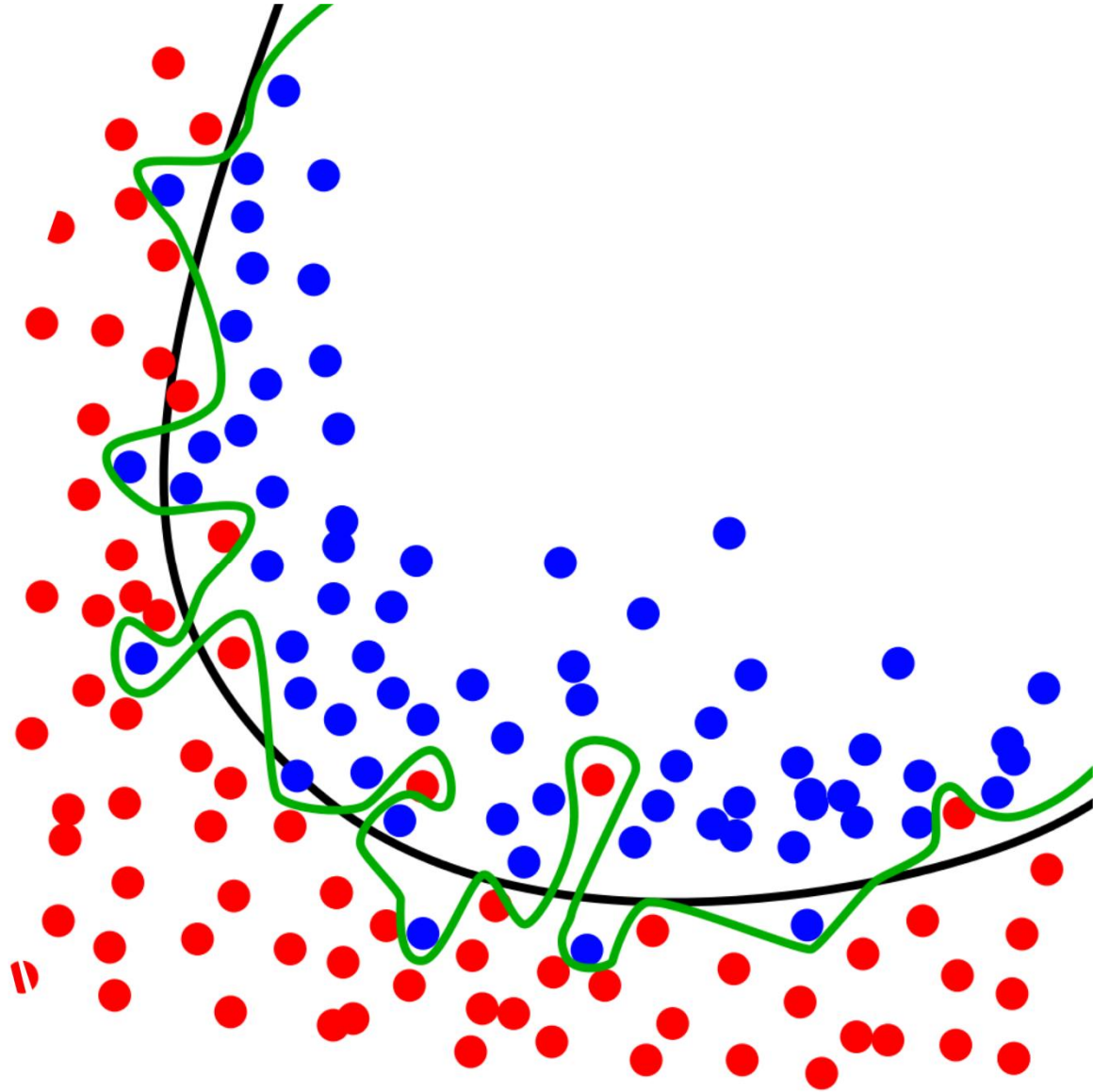


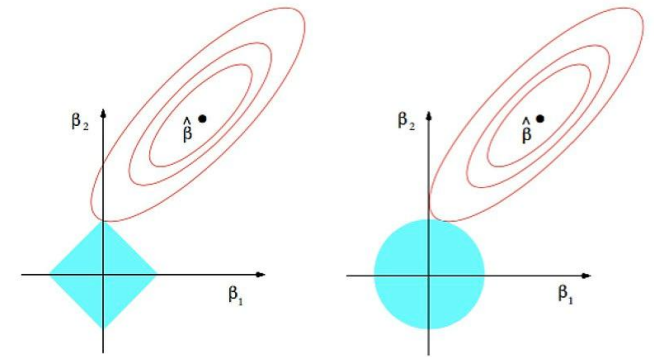
# Regularização

- Técnicas para minimizar Super Ajuste em ML
- Técnicas Específicas para RNA/DL



# Regularização

- Prevenir overfitting em ML em geral
- Um “termo” de regularização é adicionado à medida que os pesos são aprendidos
- O termo L1 é a soma dos pesos
- O termo L2 é a soma do quadrado dos pesos



Xiaoli C. [CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0>)]

# Qual é a diferença?

## L1: soma dos pesos

- Executa a seleção de atributos - atributos inteiros vão para 0
- Computacionalmente ineficiente
- Saída esparsa

## L2: soma do quadrado dos pesos

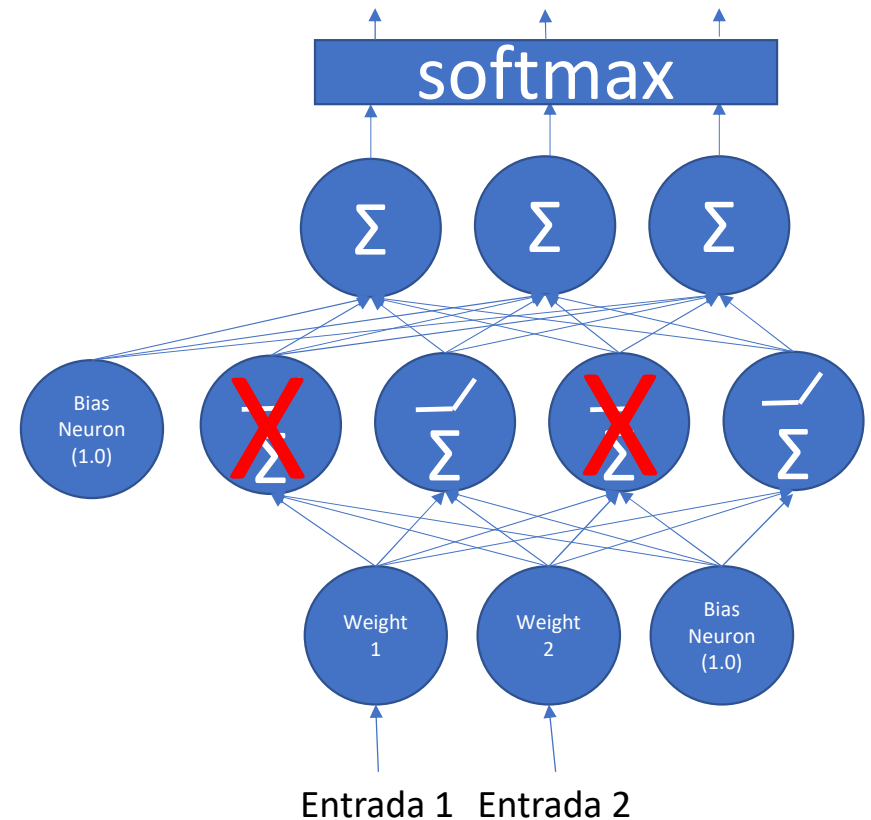
- Todos os atributos permanecem, são apenas ponderados
- Computacionalmente eficiente
- Saída densa

Por que  
você iria  
usar L1?

- A seleção de atributos pode reduzir a dimensionalidade
  - De 100 atributos, talvez apenas 10 acabem com coeficientes diferentes de zero!
  - A dispersão resultante pode compensar sua ineficiência computacional
- Mas, se você acha que todos os seus atributos são importantes, L2 é provavelmente uma escolha melhor.

# Dropout

- Remove aleatoriamente neurônios e suas conexões
- Normalmente tem uma performance melhor, pois funciona como se fossem vários modelos
- É um hiper parâmetro



# Early stopping

Train on 768 samples

Epoch 1/150

768/768 [=====] - 1s 1ms/sample - loss: 2.2462 - accuracy: 0.5417

Epoch 2/150

768/768 [=====] - 0s 160us/sample - loss: 1.6865 - accuracy: 0.5586

Epoch 3/150

768/768 [=====] - 0s 141us/sample - loss: 1.3941 - accuracy: 0.5859

Epoch 4/150

768/768 [=====] - 0s 142us/sample - loss: 1.1572 - accuracy: 0.6146

Epoch 5/150

768/768 [=====] - 0s 197us/sample - loss: 1.0420 - accuracy: 0.6393

Epoch 6/150

768/768 [=====] - 0s 191us/sample - loss: 0.9022 - accuracy: 0.6615

Epoch 7/150

768/768 [=====] - 0s 148us/sample - loss: 0.8634 - accuracy: 0.6576

Epoch 8/150

768/768 [=====] - 0s 140us/sample - loss: 0.7967 - accuracy: 0.6484

Epoch 9/150

768/768 [=====] - 0s 140us/sample - loss: 0.7984 - accuracy: 0.6732

Epoch 10/150

768/768 [=====] - 0s 140us/sample - loss: 0.7989 - accuracy: 0.6576

# Data Augmentation

- Gerar mais dados de treino
- Aplicados especialmente com imagens:
  - Girar
  - Inverter
  - Alterar tamanho



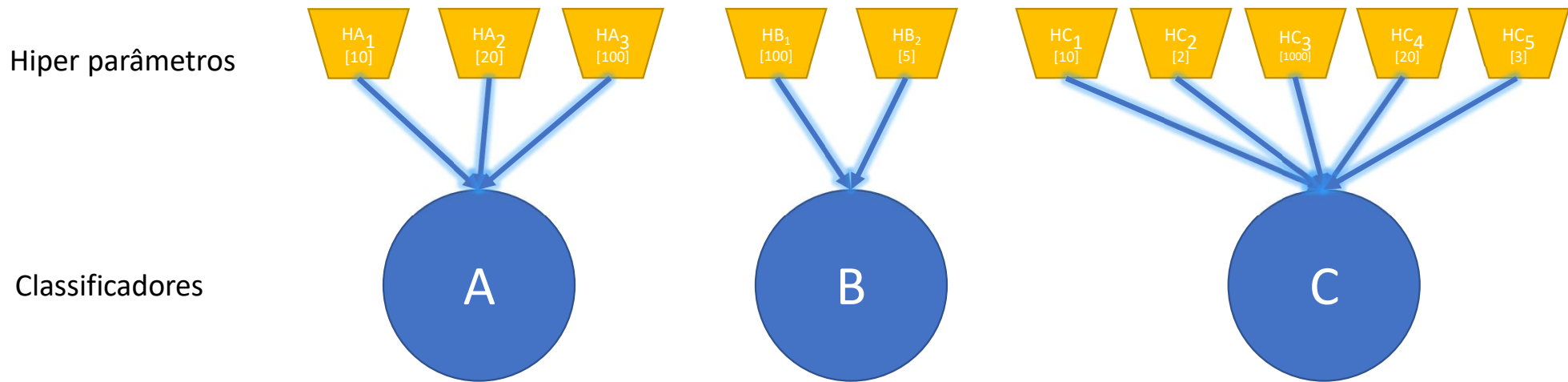
# Hiper Parâmetros

- Parâmetros: configurações que não definimos diretamente
  - Pesos das sinapses em uma RNA
- Hiper Parâmetros: configurações que definimos previamente a criação do modelo
  - Epochs
  - Batch size





# Usando mais Classificadores



Serão necessários 1.220.500 treinos diferentes para testar todas as opções  
 $(10*20*100)+(100*5)+(10*2*1000*20*3)$

Supondo que cada configuração de cada classificador leve 1 minuto para treino

**2,3 Anos!**

# Auto Machine Learning

- Técnicas para busca OTIMIZADA de hiper parâmetros que produzem o melhor modelo possível

