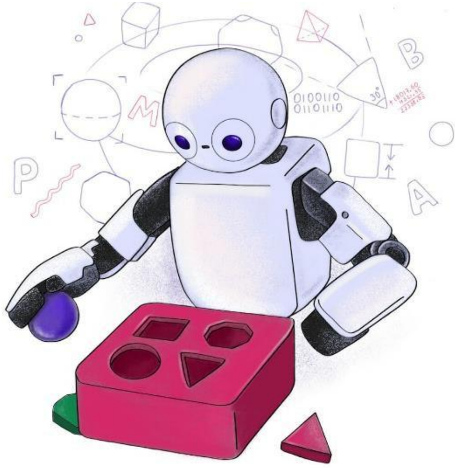
TP555 - Inteligência Artificial e Machine Learning: *Variações do formato da superfície de erro*



Esse material foi desenvolvido e gentilmente cedido pelo Prof. Dr. Felipe Augusto Pereira de Figueiredo, do Inatel.(felipe.figueiredo@inatel.br)

Inatel

Prof. Dr. Luiz Augusto Melo Pereira
luiz.melo@inatel.br

Recapitulando

- Vimos anteriormente como plotar a superfície de erro através da variação dos valores dos pesos e anotando os respectivos erros.
- No exemplo que vimos, a superfície tinha o formato de tigela, com as linhas da superfície de contorno sendo círculos.
- Isso indica que o erro varia igualmente para variações de todos os pesos.
- Agora veremos que ***nem toda superfície de erro tem formato de tigela***, em alguns casos, elas têm o formato de ***vale***.

Variações do formato da superfície de erro

- ***Nem toda superfície de erro tem formato de tigela***, em alguns casos, elas têm o formato de ***vale*** (se assemelham a um V ou U).
- Porém, ***independente do formato todas continuam sendo convexas***.
- Ou seja, continuam tendo apenas ***um ponto de mínimo***.
- Para demonstrar isso vamos supor a seguinte ***função observável***

$$y_{\text{noisy}}(n) = y(n) + w(n),$$

onde a ***função objetivo*** é dada por

$$y(n) = a_1 x_1(n) + a_2 x_2(n).$$

Variações do formato da superfície de erro

- Agora, suponhamos que nós quiséssemos aproximar a ***função objetivo*** com a seguinte ***função hipótese***

$$h(\boldsymbol{x}(n), \widehat{\boldsymbol{a}}) = \hat{y}(n) = \hat{a}_1 x_1(n) + \hat{a}_2 x_2(n).$$

- Substituindo a ***função hipótese*** na ***função de erro***, nós temos

$$J_e(\boldsymbol{a}) = \frac{1}{N} \sum_{n=0}^{N-1} \left[y_{\text{noisy}}(n) - \left(\hat{a}_1 x_1(n) + \hat{a}_2 x_2(n) \right) \right]^2.$$

- Observando a função de erro, o que você acha que ocorreria caso o intervalo de variação de x_1 fosse muito maior do que o de x_2? (ou o de x_2 ser muito maior do que o de x_1?)
  - Por exemplo, se $1000 \leq x_1 \leq 2000$ e $0 \leq x_2 \leq 1$.

Variações do formato da superfície de erro

- Caso $x_1(n) \gg x_2(n), \forall n$, então $x_1(n)$ terá uma ***influência maior no erro resultante***, o que pode ser expresso de forma aproximada como

$$J_e(\boldsymbol{a}) \approx \frac{1}{N}\sum_{n=0}^{N-1}\left[y_{\text{noisy}}(n) - \hat{a}_1 x_1(n)\right]^2.$$

- Portanto, o erro entre y_{noisy} e $h\left(\boldsymbol{x}(n)\right)$ será ***dominado pelo atributo*** $x_1(n)$ e, portanto, pequenas variações de $\hat{a}_1$ farão com que o erro varie rapidamente.
- Algo similar ocorre se $x_2(n) \gg x_1(n)$, nesse caso, o erro será ***dominado pelo atributo*** $x_2(n)$ e, portanto, pequenas variações de $\hat{a}_2$ farão com que o erro varie rapidamente.
- Vamos ver como fica o formato da superfície para estes casos.

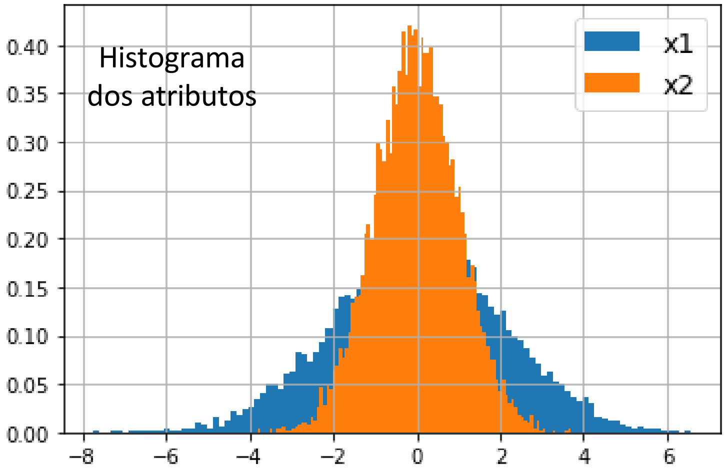
Variações do formato da superfície de erro

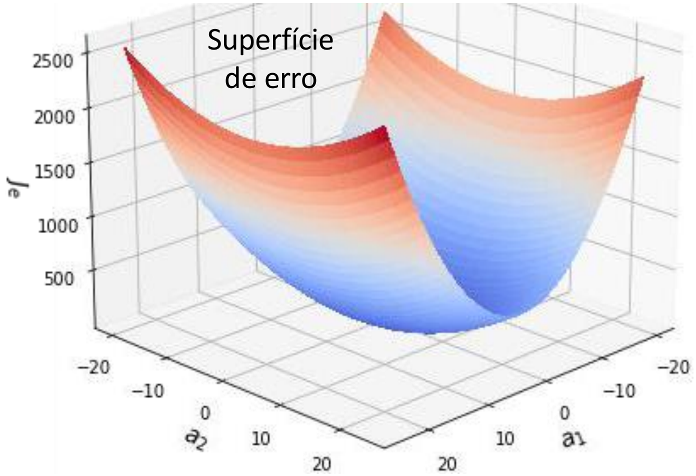
- ***Primeiro caso***: x_1 tem intervalo de variação maior do que x_2.
- Portanto, a ***influência*** da variação de $\hat{a}_1$ no ***erro*** é maior.
- Ou seja, o erro varia mais rapidamente com variações de $\hat{a}_1$, resultando em uma superfície com formato de ***vale***.
    - A superfície tem inclinação maior no sentido do peso $\hat{a}_1$.
- O erro varia bem mais lentamente com variações de $\hat{a}_2$.
- A inclinação da superfície é bem menor no sentido de $\hat{a}_2$.
- A abertura do vale está no sentido de $\hat{a}_1$.

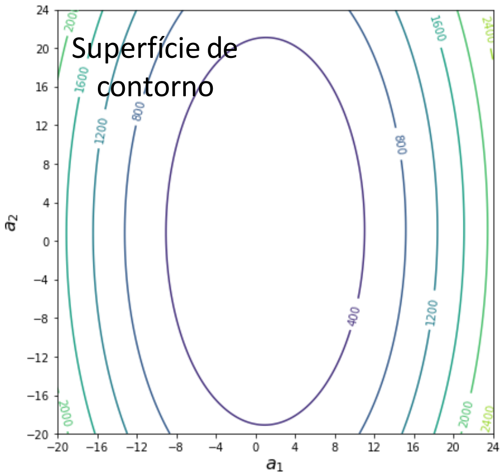
Atributos

$x_1 = 2 * \mathrm{randn}(\mathrm{N}, 1)$

$x_2 = \mathrm{randn}(\mathrm{N}, 1)$







[Exemplo: formatos_diferentes_da_superfície_de_erro.ipynb]

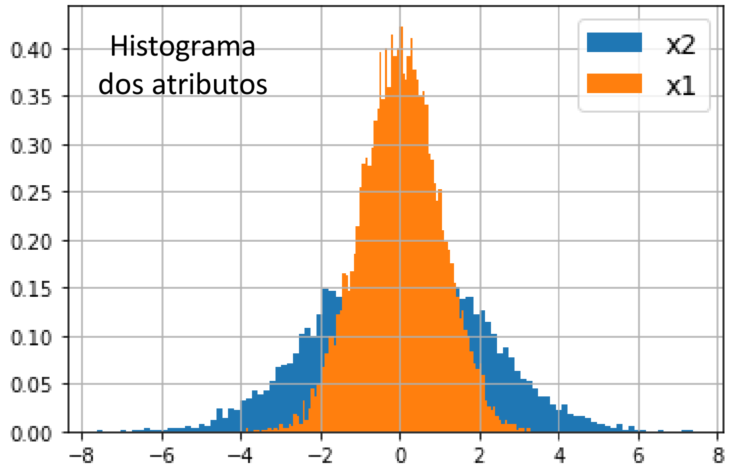
Variações do formato da superfície de erro

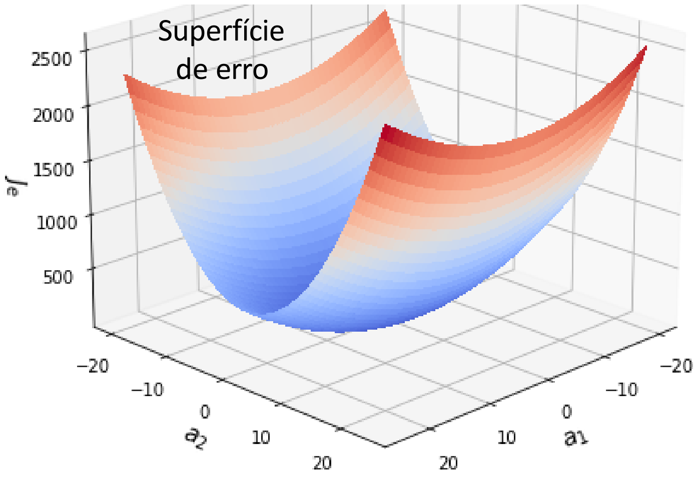
- ***Segundo caso***: x_2 tem intervalo de variação maior do que x_1.
- Então, a influência da variação de $\hat{a}_2$ no erro é maior, resultando em uma superfície com formato de ***vale***.
  - A superfície tem inclinação maior no sentido do peso $\hat{a}_2$.
- O erro varia bem mais lentamente com variações de $\hat{a}_1$.
  - A inclinação da superfície é bem menor no sentido de $\hat{a}_1$.
- A abertura do vale está no sentido de $\hat{a}_2$.

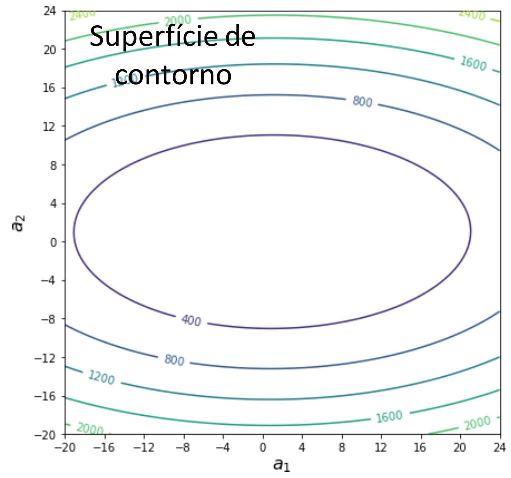
Atributos

$x_1 = \mathrm{randn}(\mathrm{N}, 1)$

$x_2 = 2 * \mathrm{randn}(\mathrm{N}, 1)$







Exemplo: formatos_diferentes_da_superfície_de_erro.ipynb

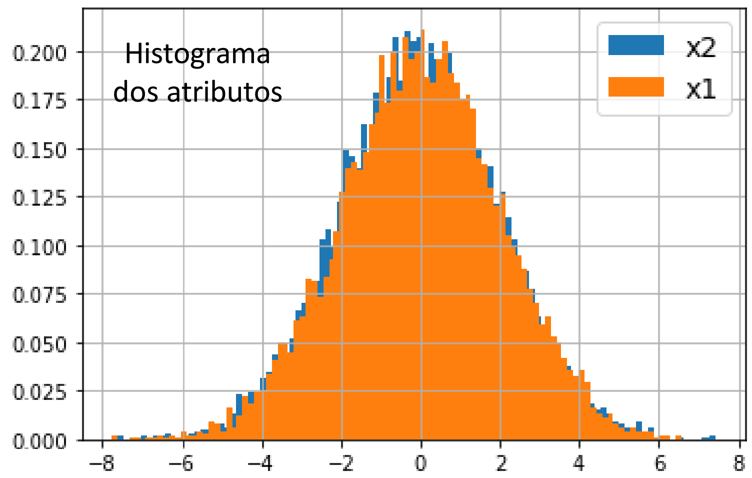
Variações do formato da superfície de erro

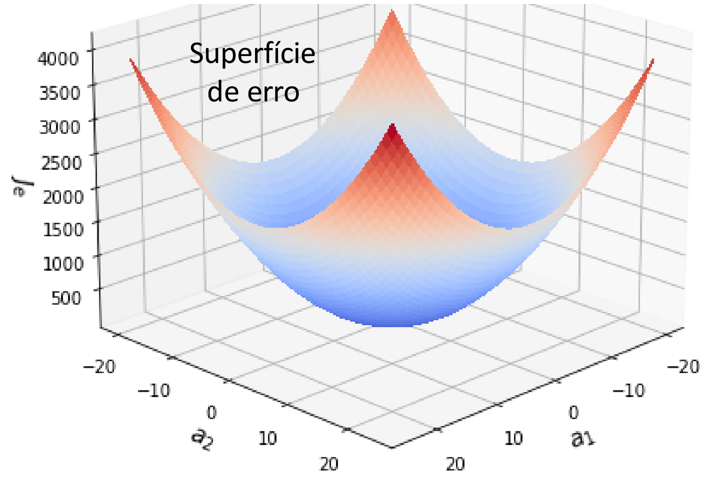
- ***Terceiro caso***: x_1 e x_2 têm intervalos semelhantes.
- Portanto, a variação tanto de $\hat{a}_1$ quanto de $\hat{a}_2$ tem influência semelhante na variação do erro, resultando em uma superfície com formato de ***tigela***.
- O erro varia de forma similar com variações de $\hat{a}_1$ ou $\hat{a}_2$.
  - A superfície tem inclinação semelhante em ambas as direções.

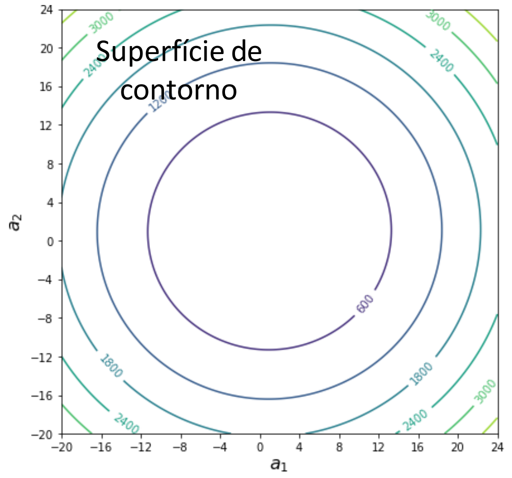
Atributos

$x_1 = 2 * \text{randn}(\text{N}, 1)$

$x_2 = 2 * \text{randn}(\text{N}, 1)$







[Exemplo: formatos_diferentes_da_superfície_de_erro.ipynb]

Variações do formato da superfície de erro

- Nós veremos em breve que superfícies com formato de vale têm uma influência negativa na convergência do algoritmo de otimização iterativo que iremos utilizar.
- Essas superfícies fazem com que a convergência seja lenta, levando muito tempo para que encontremos o melhor modelo.
- A convergência se torna lenta devido à inclinação da superfície ser diferente nas direções dos pesos.
- Em uma das direções a inclinação é grande, mas em outra ela é praticamente nula.
- Veremos também que o melhor formato para a superfície de erro é o de uma tigela, pois sua inclinação é grande e similar em todas as direções, fazendo com que a convergência seja rápida.

Obrigado!