

# Gaussian Naive bayes

Gaussian Naive bayes is named after the Gaussian distributions that represent the data in the training dataset

**Initial guesses** are called **Prior Probabilities**

**Note:** When we get really small numbers, it is a good idea to take the  $\log()$  of everything to prevent something called **underflow**.

Underflow =

Every very computer has a limit to how close a number can get to 0 before it can no longer accurately keep track of that number.

**Note:** any log will do, but the natural log ( $\ln$ ) or log base e, is the most commonly used  $\log()$  in statistic and machine learning.



$\log(\text{Loves Troll 2 Score}) = -124$



And since the score for  
**Does Not Love Troll 2 (-48)**  
is greater than the score for  
**Loves Troll 2 (-124)**...

$\log(\text{Does not Love Troll 2 Score}) = -48$

