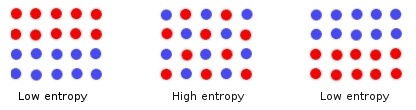
Imagine 3 buckets with 10 red and 10 blue balls as shown in the picture. The first and last have well separated (or ordered) blue and red balls, and have low entropy. The middle one has uniformly placed red and blue balls (unordered or random) and has high entropy.   
So, maximum entropy is achieved when we have uniform distribution of things or in other words when they have the most randomness.  
  
Maximum entropy classifier as the name suggest is related to maximum entropy. It is a classifier which prefers the uniformity or maximum entropy if no data is observed. But as it sees the data, it has to move away from the maximum entropy by explaining data. After it has explained the data, it again tries to maximize the entropy on whatever remaining is not seen.



The motivating idea behind maximum entropy is that one should prefer the most uniform models.

. Maximum entropy has already been widely used for a variety of natural language tasks, including language modeling

Maximum entropy has been shown to be a viable and competitive algorithm in these domains. Maximum entropy is a general technique for estimating probability distributions from data. The over-riding principle in maximum entropy is that when nothing is known, the distribution should be as uniform as possible, that is, have maximal entropy. Labeled training data is used to derive a set of constraints for the model that characterize the class-specific expectations for the distribution. Constraints are represented as expected values of “features,” any real-valued function of an example.

reference : http://www.kamalnigam.com/papers/maxent-ijcaiws99.pdf