To understand these terms, you first need to understand the concept of likelihood. Assume you have a probability distribution - or rather family of such distributions - p(x;w) which assigns a probability to each data point x, given a specific setting of its parameters w. That is, different values of the parameters, w, will change the probability assigned to each data point, x.

Now, since different parameters correspond to different distributions, we can tune the parameters in such a way that the data that we observe, D, is assigned a high probability and possible data that we don't observe is assigned a low probability. To this end we define the likelihood function L(D;w) = product\_{x in D} p(x;w). That is the likelihood is just the joint probability of the observed data as a function of the parameters.

**Maximum likelihood estimation (MLE) simply means that we seek the parameters w that maximizes the likelihood function. That is to say, we seek the parameters that assigns as much mass as possible to the observed data.**

One pitfall with the maximum likelihood estimate should be obvious from the above definition. If we assign all the probability only to the things we actually observed, we won't have any probability mass left for all the things that we didn't observe. One way to conquer this is to add a prior distribution over the parameters w, such as a Gaussian in the case of logistic regression or a Dirichlet prior in the case of the multinomial distribution, which allows us to control how probability mass should be assigned to observed and unobserved data. So how can we make use of such a prior distribution? By means of Bayes rule, we can derive the a posteriori distribution over the parameters, conditioned on the data: p(w;D) = p(D;w)\*p(w) / Z, where Z is a normalization factor which assures that p(w;D) sums to 1.

**Maximum a posteriori (MAP) estimation means that we seek the parameters w that maximize the posterior distribution p(w;D). Since Z is a constant factor, it does not affect the choice of w that maximize the posterior and we can therefore simply find the w that maximize p(D;w)\*p(w).**

Both MLE and MAP are point estimation methods, they only return a single value of the parameters w. This means that any information on the uncertainty of the parameters are lost, which is unfortunate since knowledge about this uncertainty can be used to compute things like confidence in our predictions. In order to keep this uncertainty we may adopt a fully Bayesian approach, in which we instead of finding the MLE or MAP, we find the full posterior distribution p(w;D). In this case we need to compute the normalization factor Z, which might be difficult or even intractable depending on the structure of p(x;w). This is one of the reasons why point estimates are popular.

Finally, Expectation Maximization (EM) is a technique to cope with parts of the model that we cannot observe, but we assume should be there. These unobserved parts can be either data that is simply missing from the data, but it might also be things that are never observed, but whose relation to the observed parts we can model (or in other words, parts that "explain" the observed data). In topic models, for example, the hidden parts are the topics that are assumed to explain the observed words. Assume that for each observed point x (e.g. a vector of words), we have a hidden variabel z (e.g. a topic). We then have a distribution p(x,z;w) which assigns a probability to x and z jointly. EM is a method for maximizing the joint likelihood of both the observed and hidden parts. This is done by iteratively using the current value of the parameters to estimate a distribution over the hidden variables given the observed part (expectation) and then finding the parameters that maximize the joint likelihood of the observed variables and the setting of the hidden variables from the previous step (maximization).

If you want to get a more solid understanding of these topics, I suggest you read Christopher Bishop's book. [Pattern Recognition and Machine Learning](http://research.microsoft.com/en-us/um/people/cmbishop/prml/).