Maximum A Posteriori Estimation

another blocks for inferring hidden parameters

MLE is a point estimation, that is , the form for the distribution of totality is explicitly given, but with unknown parameters, with the help of some examples sampled from the totality, we estimate the value of parameters in probability distribution. But It didn't take the distribution of parameter θ itself into account. i.e., Suppose $\theta \in \mathbf{\Theta}$, In MLE, this distribution is uniform distribution, the probability is fixed in this case. While sometimes before observing the data, we have a prior knowledge giving the distribution of parameter θ, $\mathbf{p}(\theta)$

Bayesian estimation:

$$\mathbf{p}(\theta|\mathbf{x}^1, \ldots, \mathbf{x}^m) = \frac{\mathbf{p}(\mathbf{x}^1, \ldots, \mathbf{x}^m|\theta)\mathbf{p}(\theta)}{\mathbf{p}(\mathbf{x}^1, \ldots, \mathbf{x}^m)}$$

Deep Learning p135

In the scenarios where Bayesian estimation is typically used, the Prior $\mathbf{p}(\theta)$ begins as a relatively uniform or Gaussian distribution with high entropy. and the m observations of the data ususlly causes the posterior to lose entropy and concentrate around a few highly likely values of the parameters.