Foundations Of Data Science

Assignment-3 Report

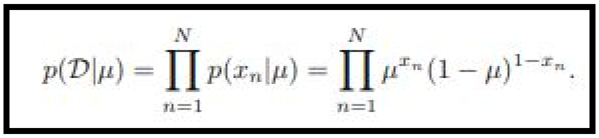
ID Number	Name
2017A7PS0121H	Prakhar Gupta

Part A: Sequential Learning

Posterior distribution ∝ Likelihood distribution x Prior distribution
P (μ |D, a, b) ∝ P (D |μ) x P (μ |a, b)

Coin-tossing follows a Bernoulli distribution. It’s probability density function is given by
Bern (x |μ) = (μ^x)*(1−μ)^(1−x)

where μ is the mean of the Bernoulli distribution.
Thus, for a dataset D of N points, we get the likelihood function as



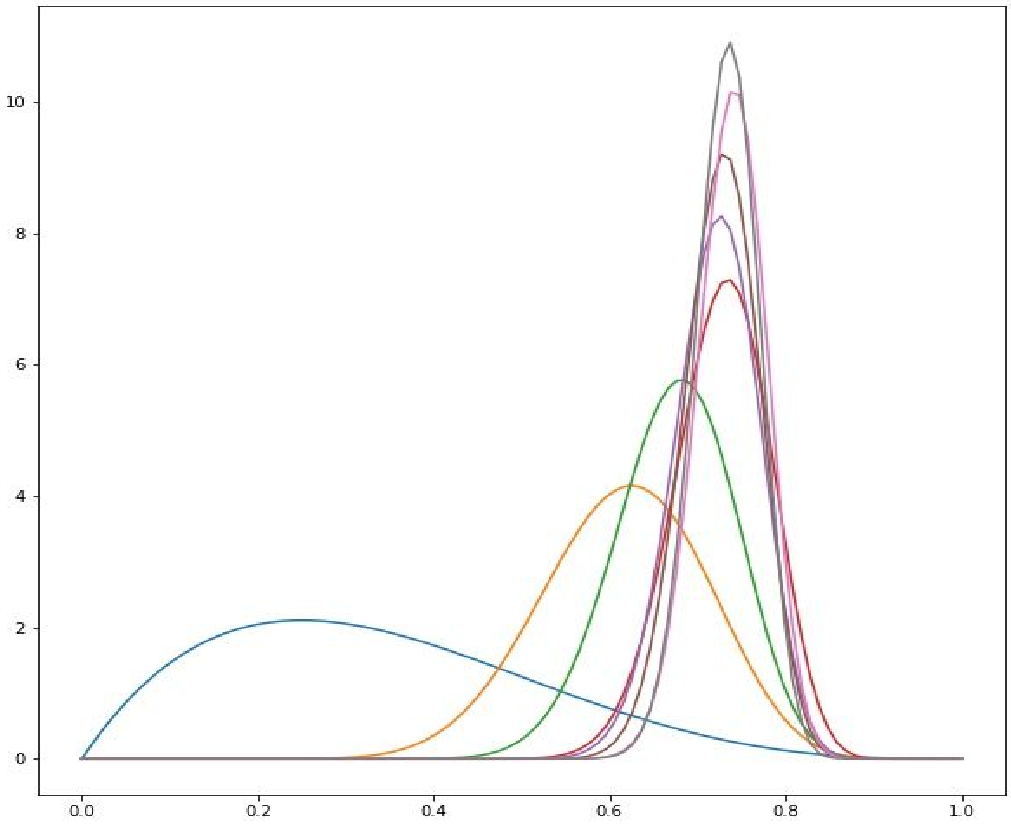
$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}.$$

We will take the prior to be a beta distribution. The PDF for a beta distribution is given by

$$\boxed{\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}}$$

where a and b are the parameters and Γ is the gamma function.

Sequential learning plot : PDF For Posterior (beta) Distribution



In sequential learning, one data point will be taken at a time and hence the likelihood function will have only one term in its product

As with any sequential learning problem, this posterior after viewing the first example becomes the prior for the next example. Thus, we start off with a prior and will observe how examples coming in change the appearance of the prior.

Part B: Combined Learning

(using all points at once for generating posterior distribution)

Posterior distribution ∝ Likelihood distribution x Prior distribution
P (μ |D, a, b) ∝ P (D |μ) x P (μ |a, b)

Coin-tossing follows a Bernoulli distribution. It’s probability density function is given by
Bern (x |μ) = (μ^x)*(1−μ)^(1−x)

where μ is the mean of the Bernoulli distribution.
Thus, for a dataset D of N points, we get the likelihood function as

$$\boxed{p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}.}$$

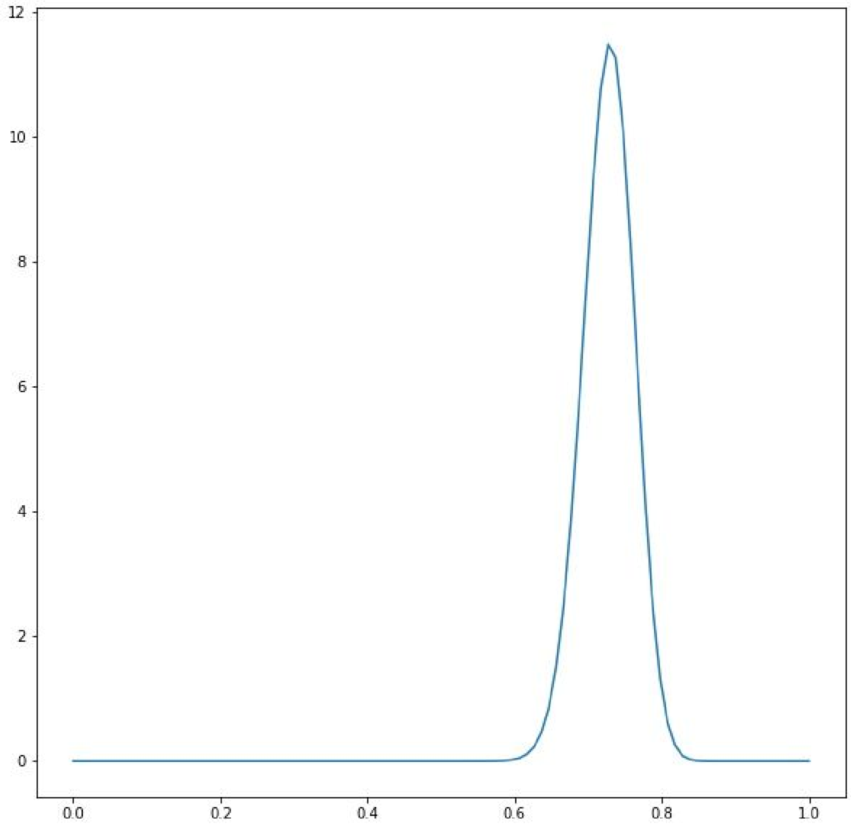
We will take the prior to be a beta distribution. The PDF for a beta distribution is given by

$$\boxed{\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}}$$

where a and b are the parameters and Γ is the gamma function

In this case, the likelihood will be over the entire dataset and hence we compute the posterior after viewing the entire dataset at once.

Combined learning plot : PDF For Posterior(beta) Distribution



Part C: Comments/Inferences

We see that this sequential approach to learning arises naturally when we adopt a Bayesian viewpoint. It is independent of the choice of prior and of the likelihood function and depends only on the assumption of the data. Sequential methods make use of observations one at a time, or in small batches, and then discard them before the next observations are used.

They can be used, for example, in real-time learning scenarios where a steady stream of data is arriving, and predictions must be made before all of the data is seen. Because they do not require the whole data set to be stored or loaded into memory, sequential methods are also useful for large data sets. Sequential models tend to learn and adapt after each data point.

Whereas in the case of the second model the posterior distribution is generated based on cumulative learning where all data points at once are used to generate the maximum likelihood estimator and hence posterior distribution based on maximum likelihood estimator generated.

Both sequential and combined learning models finally give the same posterior curve on the same number of data points. Also in both the models, we get the same the final type of posterior distribution.

From the sequential learning plot, we see that as the number of observations increases or if we increase data points, so the posterior distribution becomes more sharply peaked.

This can also be seen from the result for the variance of the beta distribution, in which we see that the variance goes to zero for $a \to \infty$ or $b \to \infty$. As we increase data points posterior distribution reaches our posterior curve comes closer to pdf for the underlying distribution or prior distribution (of data points).

If $\mu ML = 0.5$ then the coin becomes unbiased and the posterior distribution becomes a steep curve symmetric about $x=0.5$.