CMPT 423/820

Assignment 1 Question 8

- Seyedeh Mina Mousavifar
- 11279515
- sem311

Task

Derive $P(y|\mathbf{X}_1, \mathbf{X}_2) = \frac{m_1 + m_2 + a}{N_1 + N_2 + a + b}$

Based on *Conditional Probability* formula we have:

$$P(\mu|\mathbf{X}_1, \mathbf{X}_2) = \frac{P(\mu, \mathbf{X}_1, \mathbf{X}_2)}{P(\mathbf{X}_1, \mathbf{X}_2)}$$

Because of *Commutative property of Intersection* we can rewrite the equation as:

$$P(\mu|\mathbf{X}_1, \mathbf{X}_2) = \frac{P(\mathbf{X}_2, \mu, \mathbf{X}_1)}{P(\mathbf{X}_1, \mathbf{X}_2)}$$

In order to obtain yesterday knowledge, we apply *Bayes' Rule* for $P(\mathbf{X}_2, \mu, \mathbf{X}_1)$:

$$P(\mu|\mathbf{X}_1, \mathbf{X}_2) = \frac{P(\mathbf{X}_2|\mu, \mathbf{X}_1)\, P(\mu, \mathbf{X}_1)}{P(\mathbf{X}_1, \mathbf{X}_2)}$$

Applying *Conditional Probability* formula for $P(\mu, \mathbf{X}_1)$:

$$P(\mu|\mathbf{X}_1, \mathbf{X}_2) = \frac{P(\mathbf{X}_2|\mu, \mathbf{X}_1)\, P(\mu|\mathbf{X}_1)\, P(\mathbf{X}_1)}{P(\mathbf{X}_1, \mathbf{X}_2)}$$

As we can see, we have obtained our prior knowledge, which is $P(\mu|\mathbf{X}_1)$.

If we apply *Conditional Probability* formula for $P(\mathbf{X}_2|\mathbf{X}_1)$, we will have:

$$P(\mu|\mathbf{X}_1, \mathbf{X}_2) = \frac{P(\mathbf{X}_2|\mu, \mathbf{X}_1)\, P(\mu|\mathbf{X}_1)}{\frac{P(\mathbf{X}_1, \mathbf{X}_2)}{P(\mathbf{X}_1)}} = \frac{P(\mathbf{X}_2|\mu, \mathbf{X}_1)\, P(\mu|\mathbf{X}_1)}{P(\mathbf{X}_2|\mathbf{X}_1)}$$

Considering the term $P(\mu, \mathbf{X}_1)$, we have already estimated μ based on $\mathbf{X}_1$, so we can infer that μ convey $\mathbf{X}_1$ in itself, and we can omit $\mathbf{X}_1$.

$$P(\mu|\mathbf{X}_1, \mathbf{X}_2) = \frac{P(\mathbf{X}_2|\mu)\, P(\mu|\mathbf{X}_1)}{P(\mathbf{X}_2|\mathbf{X}_1)}$$

As we can see in the above equation, likelihood is $P(\mathbf{X}_2|\mu)$, prior probability is $P(\mu|\mathbf{X}_1)$, and $P(\mathbf{X}_2|\mathbf{X}_1)$ is the normalization constant. Besides:

$$P(\mathbf{X}_2|\mu) = Bin(m_2|N_2, \mu) \quad \textit{Binomial distribution}$$
$$P(\mu|\mathbf{X}_1) = Beta((\mu|\mathbf{X}_1)|a', b') \quad \textit{Beta distribution}$$

Above, we can see that our knowledge of yesterday, has changed our prior Beta distribution. So, we will be more biased toward yesterday data, and add that info to our prior bias. So:

$$P(\mu|\mathbf{X}_1) = Beta((\mu|\mathbf{X}_1)|a', b') \quad a' = a + m_1 , \quad b' = b + N_1 - m_1$$
$$\Rightarrow P(\mu|\mathbf{X}_1) = Beta((\mu|\mathbf{X}_1)|a + m_1, b + N_1 - m_1)$$

So for finding the final result, we have:

$$P(\mu|\mathbf{X}_1, \mathbf{X}_2) = \frac{P(\mathbf{X}_2|\mu)\, P(\mu|\mathbf{X}_1)}{P(\mathbf{X}_2|\mathbf{X}_1)}$$

$$\Rightarrow P(y|\mathbf{X}_1, \mathbf{X}_2) = E[\mu|\mathbf{X}_1, \mathbf{X}_2] = \int_\mu \mu \frac{P(\mathbf{X}_2|\mu)\, P(\mu|\mathbf{X}_1)}{P(\mathbf{X}_2|\mathbf{X}_1)} d\mu$$

We have seen that:

$$P(y|\mathbf{X}) = \int_\mu \mu \frac{P(\mathbf{X})P(\mu)}{P(\mathbf{X})} d\mu = \frac{m + a}{N + a + b}$$

So we infer that the integral, would give the same result, only with new parameters, which were discussed above:

$$P(y|\mathbf{X}_1, \mathbf{X}_2) = E[\mu|\mathbf{X}_1, \mathbf{X}_2] = \int_\mu \mu \frac{P(\mathbf{X}_2|\mu)\, P(\mu|\mathbf{X}_1)}{P(\mathbf{X}_2|\mathbf{X}_1)} d\mu = \frac{m_2 + a'}{N_2 + a' + b'}$$