Bayesian Machine Learning, MLE, MAP - II

Nipun Batra

February 8, 2020

IIT Gandhinagar

Fully Bayesian Approach

- MLE and MAP do not give us uncertainty.

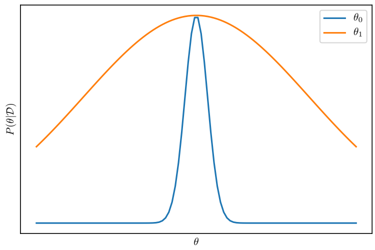

Figure 1: More uncertainty around θ_0 than θ_1

Predictive Distribution for the Coin Toss Problem

- $P(\text{Next Toss} = H | \text{ Data})$
- What value of θ should be used?
- Answer: Use all the possible values of θ.
- $P(\text{Next Toss} = H|Data) = \int P(\text{Next Toss} = H, \theta| \text{ Data})d\theta$
  Why?
- Marginalization! $p(x) = \int_y p(x, y)dy$
- Note that the conditioning above is only over the observed data or evidence.

Predictive Distribution for the Coin Toss Problem if θ is known

- Let c be a random variable that is assigned the value 1 if head results after tossing a coin and 0 if tail results after tossing a coin.
- Question: What is $P(\text{Next Toss} = c|\theta)$?
- Answer: $\theta^c(1-\theta)^{1-c}$. Why?
- Suppose c = 0. Then $P(\text{Tails}|\theta) = (1-\theta)$.

Predictive Distribution for the Coin Toss Problem

- Let us consider the case where we have a Beta prior for our coin toss problem. What is the predictive distribution, given we have observed some data?
- Answer: $P(\text{Next} = c|\mathcal{D}, a, b) = \int P(\text{Next} = c, \theta|\mathcal{D}, a, b)d\theta$
- From the chain rule of probability, we have the following:

$$P(AB|CDE) = \frac{P(ABCDE)}{P(CDE)} = \frac{P(A|BCDE)P(BCDE)}{P(CDE)}$$

$$= P(A|BCDE)P(B|CDE)$$

Predictive Distribution for the Coin Toss Problem

- In our case, the integrand $P(\text{Next} = c, \theta | \mathcal{D}, a, b)$ therefore becomes, $P(\text{Next} = c, | \theta, \mathcal{D}, a, b) P(\theta | \mathcal{D}, a, b)$
- If θ is known, then $P(\text{Next} = c | \theta, \mathcal{D}, a, b) = P(\text{Next} = c | \theta)$. Why?
- This is because we know the actual model parameter distribution. The data *cannot* affect it. What about a and b affecting the prior? They do not concern us anymore either, since we actually know the parameters.
- $\implies$ The predictive distribution is,

$$\int_\theta \theta^c (1-\theta)^{1-c} \frac{\Gamma(n_H + n_T + a + b) \theta^{n_H + a - 1} (1-\theta)^{n_T + b - 1} d\theta}{\Gamma(n_H + a) \Gamma(n_T + b)}$$

Predictive Distribution for the Coin Toss Problem

$$= \frac{\Gamma(n_H + n_T + a + b)}{\Gamma(n_H + a)\Gamma(n_T + b)} \int_\theta \theta^{n+h+a-1+c}(1-\theta)^{n_T+b-c} d\theta$$

$$= \frac{\Gamma(n_H + n_T + a + B)\Gamma(c + n_H + a)\Gamma(n_T + b - c + 1)}{\Gamma(n_H + a)\Gamma(n_T + b)\Gamma(1 + n_H + a + n_T + b)}$$