BITS F464: Machine Learning

05 Towards Bayesian Machine Learning

Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

Jan 27, 2021 ONLINE (Campus @ BITS-Pilani Jan-May 2021)

http://ktiwari.in/ml

Let's change our focus a bit

- Input: x Output: y
- Training data: $(x^{(1)}, y^{(1)})$, $(x^{(2)}, y^{(2)})$, ..., $(x^{(m)}, y^{(m)})$
- Target function $f : x \rightarrow y$
- Hypothesis $h : x \rightarrow y$

Instead of reporting y let's report $P(y)$

- For a given input x, output is not True/False
- But,

$$P(True) = 0.3$$
$$P(False) = 0.7$$

Bayes Theorem

Example: Three companies A, B and C makes 35%, 35% and 30% of all the lamps in market. Probability of their lamp being defective is 1.5%, 1% and 2% respectively. What is the probability that a randomly selected defective lamp was manufactured in factory C?

Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Prior, posterior, likelihood and evidence ($P(A)$, $P(A|B)$, $P(B|A)$, $P(B)$)

- What we want is $P(C|D)$
- $P(D) = P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) = .35 * 0.015 + .35 * 0.01 + .3 * 0.02 = 0.01475$

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} = \frac{0.02 * 0.3}{0.01475} = 0.407$$

Let's Predict

Consider table below: Can you determine the value of ???

x1	x2	x3	x4	x5	Y
32	45	13	39	92	0
82	70	77	35	93	1
14	50	95	98	93	0
37	23	92	39	82	0
22	18	96	47	36	1
13	70	0	31	45	1
18	87	56	49	35	0
34	2	7	41	76	1
82	4	98	20	87	0
50	14	94	22	32	0
10	39	74	69	58	0
75	53	80	6	64	1
61	30	47	37	59	1
43	67	55	7	59	0
32	87	16	8	92	1
93	63	38	1	60	0
64	22	41	15	75	1
41	51	16	11	8	???

To do this

- One have to determine how Y depends upon attributes $x1$, $x2$, $x3$, $x4$, $x5$

  $y = f(x1, x2, x3, x4, x5)$
- What about

  $f = mod(\sum w_i \times x_i, 2)$

  with $(w1, w2, w3, w4, w5) = (0, 1, 0, 1, 0)$

Bias: How to come up with these values

- Observe the data
- Assume data for 10 coin tosses be -H-H-T-H-H-T-T-H-H-H-
- Now again the coin is to be flipped, what is expected output?
- $P(H) = 0.7$ and $P(T) = 0.3$

You can incorporate your bias

Assume you know the coin was unbiased (and you have a high confidence on this)

- You hypothetically consider 100 more flips [1]. Being a fair coin it would give 50 H and 50 T
- So your estimate for probability of head is as below $P(H) = (7 + 50)/(110) = 0.518$ and $P(T) = 0.482$
- Why 100? why not 10000? OK; if you have more confidence on bias take 10000.

[1] Also called laplacian smoothing

Probability of observing a dataset

Assume you are flipping a biased coin where $p(H) = 0.4$. What is the probability that you see this dataset $D =< H, H, T, T, H, H >$

- $p(H) = 0.4$
- $p(T) = 1 - p(H) = 1 - 0.4 = 0.6$
- If **all the trails are independent** then $p(D|\theta)$

$$= p(H) \times p(H) \times p(T) \times p(T) \times p(H) \times p(H)$$
$$= 0.4^4 \times 0.6^2 = 0.009216$$

Note: Order of elements in the data set do not matter in the trial. So $p(< H, H, H, H, T, T >)$ is same (in fact any other permutation)

What is θ

It is the parameter. For our case it represents $p(H) = 0.4$

Thank You!

Thank you very much for your attention!

Queries ?