DATE: 27 10 2019

Logistic Regression in Bayes theorem Pov:

D- given data constraints.

H - Hypothesis.

Bayes theorem

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)}.$$

$P(H)$ - prior probability.

$P(D|H)$ - likelihood.

$P(H|D)$ - posterior probability.

posterior probability $\propto$ likelihood $\times$ prior prob.

$P(D)$ is something challenging to compute

$P(\bar{H}|D)$ = probability of H not happening given D.

$$\therefore \frac{P(H|D)}{P(\bar{H}|D)} = \frac{P(D|H) \cdot P(H)}{P(D|\bar{H}) \cdot P(\bar{H})}$$

DATE: 27 10 2019

Sum of posterior odds

$P(H|D) + P(\bar{H}|D) = 1$

the same can be written as

$P(H_1|D) + P(\bar{H}_1|D) = 1$

where H_1 is one of the hypothesis

$\bar{H}_1 = H_2 + H_2 + H_3 \ldots H_k$

$\therefore P(H_1|D) + P(H_2|D) + P(H_3|D) \ldots P(H_k|D) = 1$

$\frac{P(H|D)}{P(\bar{H}|D)} = \text{odds.} = O(H|D)$

intuition: if odds = 10 ⇒ 10 times likely for our H to be success compared to H being a failure.

can we explain odds ratio?

By machine learning?

when $P(H|D) \to 0$, $O(H|D) = 0$

when $P(H|D) \to 1$, $O(H|D) = \infty$

DATE: 27 10 2019

so applying log transformation on both sides.

$$\log\left[\frac{P(H/D)}{P(\bar{H}/D)}\right] = \log\left[\frac{P[D/H]}{P[D/\bar{H}]}\right] + \log\left[\frac{P(H)}{P(\bar{H})}\right]$$

$$\log(O(H/D)) = \log\left(\frac{P(D/H)}{P(D/\bar{H})}\right) + \log(O(H))$$

$\log(O(H))$ = log of odds of prior

It doesn't depend on Data (D or X)

$\therefore$ $\log(O(H))$ can be learned as β_0

$\log\left[\frac{P(D/H)}{P(D/\bar{H})}\right]$ = log of likelihood ratio.

which can be approximated as a function of D $= \beta D$.

$\therefore$ $\log(O(H/D)) = \beta D + \beta_0$

taking exponent on both sides

$O(H/D) = \exp(\beta D + \beta_0)$

DATE: 27 10 2019

$$\frac{p(H|D)}{[1-p(H|D)]} = exp(\beta D + \beta_0)$$

$$\frac{p(H|D)}{[1-p(H|D)]} = exp(\beta D + \beta_0)$$

$$p(H|D) = exp(\beta D + \beta_0) - p(H|D) \, exp(\beta D + \beta_0)$$

$$p(H|D) + p(H|D) \, exp(\beta D + \beta_0) = exp(\beta D + \beta_0)$$

$$p(H|D) = \frac{exp(\beta D + \beta_0)}{1 + exp(\beta D + \beta_0)}$$

this is analogous to the below

$$p(x) = \frac{exp(\beta_0 + \beta_1 x)}{1 + exp(\beta_0 + \beta_1 x)}$$

$$\frac{p(x)}{1 - p(x)} = exp(\beta_0 + \beta_1 x)$$