Super Agi Assignment

BRAJ RAJ NAGAR
Entry No- 2022AIB2CBL
[IIT Delhi]

[1] In logistic Regression hypothesis is given as

$$h_w(x) = \frac{1}{1+e^{w^Tx}}$$

Let $y = w^Tx = w_0^T + w_1^T x_1 + w_2^T x_2 + \cdots + w_n^T x_n$ —— (1)

if x_n is duplicated and made into new feature $x_{(n+1)}$

then

$$y = w_{new_0}^T + w_{new_1}^T x_1 + w_{new_2}^T x_2 + \cdots + w_{new_n}^T x_n + w_{new_{n+1}}^T x_{n+1}$$

But $x_n = x_{n+1}$ (duplicated feature) so

$$y = w_{new_0}^T + w_{new_1}^T x_1 + \cdots + (w_{new_n}^T + w_{new_{n+1}}^T) x_n \quad (2)$$

Comparing eqn (1) with eqn (2) we can say.

$w_{new_0} = w_0$, $w_{new_1} = w_1$, ... , $w_{new,n-1} = w_{n-1}$

and $w_n^T = w_{new,n}^T + w_{new,n+1}^T$

[2] 3. Both D and E better than A with 95% confidence.
Both B and C are worse than A with over 95% confidence.
Reason: CTR is key indicator of the effectiveness of an email campain. The higher the CTR, the more the successful the campain is.

[3] ⇒ Log likely hood ~~of~~ for logistic Regression

$$l(\theta) = \log(L(\theta))$$

$$= \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1-y^{(i)}) \log(1-h(x^{(i)}))$$

∴ gradient decent update rule

$$\theta := \theta + \alpha \nabla_\theta l(\theta)$$

$$\frac{\partial}{\partial \theta_j} l(\theta) = (y - h_\theta(x^{(i)})) x_j^{(i)}$$

so $\theta_j := \theta_j + \alpha (y^{(i)} - h_\theta(x^{(i)})) x_j^{(i)}$

$n \rightarrow$ features; $m \rightarrow$ training examples

Then computation cost will be $O(nm)$

[4] ⇒ The 3 approches to generate additional training data for the V2 Classifier and discuss their potential impact on the accuracy of the classifier

1. this approch aim to find examples that are challenging for the V1 classifier, as they are close to the decision boundry. However these examples might be ambiguous and not naccessarily represent clear cases of information or entertainment. it could introduce noise ~~in~~ into the training data.

2. this approch involve randomly selected 10,000 labeled Stories from the 1000 news source it provide diverse examples from different sources but the randomness might not gurenty a good representation of challenging cases for classification it depends on ~~on~~ the quality of the orginal labelling.

3. label 1 million stories and ~~select~~ Select CB and wrong:
label 1million stories and select 10,000
This approch labels a large datset and select 10,000 examples where the V1 classifier is both wrong and farthest away from the decision boundry. This method focuses on correct mistakes made by V1 classifier and identifies examples that are confidently misclassified.

[5] ⇒ $p \to$ head
$(1-p) \to$ Tail

(a) Likelihood $L = {}^nC_k \, p^k (1-p)^{n-k}$

$$\frac{\partial L}{\partial p} = {}^nC_k \, k p^{k-1} (1-p)^{n-k} - {}^nC_k \, p^k (n-k)(1-p)^{n-k-1} = 0$$

$$k - kp - (n-k)p = 0$$

$$\boxed{p = \frac{k}{n}}$$

(b) given prior $p(p) = \begin{cases} 1 & 0 \le p \le 1 \\ 0 & \text{otherwise} \end{cases}$

⇒ Likelihood $p(k,n) = {}^nC_k \, p^k (1-p)^{n-k}$

$$\because \; p(p \mid x=k) \propto p(x=k) \, p(p)$$

$$E(p) \Rightarrow \int_0^1 p \, {}^nC_k \, p^k \, p^{(n-k)} \, dp$$

$$E(p) = \frac{\Gamma(k+2) \, \Gamma(k+1-n)}{\Gamma(n+3)} \; {}^nC_k$$

$$\cancel{E[p] = \frac{b!}{(b-a)!\,a!} \times \frac{(a+2)!\,(b-a+1)!}{(b+3)!}}$$

$$E[p] = \frac{(n-k+1)(k+1)(k+2)}{(n+1)(n+2)(n+3)}$$

(c) $\max\left(P(p \mid n,k)\right)$

$$P(p \mid n,k) = {}^{n}C_{k}\, p^{k} (1-p)^{n-k}$$

max' at $p = \frac{k}{n}$ (from part (a))

$$P_{MAP} = \frac{k}{n}$$