Homework 1

1] Let x_c denote probability that the person is criminal.

Let x_{TA} denote probability that the person has blood of type A (rare blood type).

$$P(x_c | x_{TA}) = \frac{P(x_c, x_{TA})}{P(x_{TA})}$$

$$= \frac{P(x_{TA} | x_c) \cdot P(x_c)}{P(x_{TA})} \quad \ldots \text{①}$$

Since the blood found at scene was rare blood type, probability that person has rare blood type given that the person is criminal is 1.

$$P(x_c) = \frac{1}{y} \quad \ldots \text{(where } y \text{ is the population)}$$

$$P(x_{TA}) = 0.01 \quad \ldots \text{(given)}$$

$$\therefore P(x_c | x_{TA}) = \frac{1 \cdot 1/y}{0.01} \quad \ldots \text{(from eq}^n \text{ ①)}$$

$$P(x_c | x_{TA}) = \frac{100}{y} \quad \ldots \text{②}$$

From equation 2,

the probability that the person with crime blood type is guilty is dependent on the population size and is 1 for $y = 100$ (because 1% of populat is 1, hence the person is definitely a crimina) an varies (reduces) as the population size increases.

2] a] (was discussed in class)

$$p(D|\theta) = \theta^{N_1}(1-\theta)^{N_0} \quad \dots \quad (1)$$

To maximise $p(D|\theta)$,

differentiating both sides of equation 1,

$$\frac{d}{d\theta} p(D|\theta) = \frac{d}{d\theta} \theta^{N_1}(1-\theta)^{N_0}$$

maximum $p(D|\theta)$ is when $\frac{d}{d\theta} p(D|\theta) = 0$

Since $\log_e$ is an increasing funtion maximising $\ln p(D|\theta)$ should give the same results.

$$\frac{d}{d\theta} \ln p(D|\theta) = \frac{d}{d\theta} \ln \left(\theta^{N_1} (1-\theta)^{N_0}\right)$$

$$= \frac{d}{d\theta}\left(N_1 \ln \theta + N_0 \ln(1-\theta)\right)$$

$$= \frac{N_1}{\theta} + \frac{N_0 (-1)}{1-\theta}$$

Now

$$\frac{N_1}{\theta} + \frac{N_0(-1)}{1-\theta} = 0$$

$$\Rightarrow N_1(1-\theta) - N_0\theta = 0.$$

$$\Rightarrow N_1 = (N_1+N_0)\theta$$

$$\Rightarrow \theta = \frac{N_1}{N_1+N_0}.$$

Camlin Page

Date / /

b] $\arg\max_{\theta} p(\theta|D)$ $\rightarrow$ (To find)

$$p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{p(D)}$$

$$= \frac{\theta^{N_1}(1-\theta)^{N_0} \cdot k\,\theta^{\alpha}(1-\theta)^{\alpha}}{P(D)}$$ .. [from equation ① and ④ in assignment question

$$= \frac{k\,\theta^{N_1+\alpha}(1-\theta)^{N_0+\alpha}}{P(D)}$$

Since $\frac{k}{P(D)}$ is independent of θ,

$$\arg\max \frac{k\,\theta^{N_1+\alpha}(1-\theta)^{N_0+\alpha}}{P(D)} = \arg\max \underline{\theta^{N_1+\alpha}(1-\theta)^{N_0+\alpha}}$$

For maximising $\theta^{N_1+\alpha}(1-\theta)^{N_0+\alpha}$

following method similar in part A,

$$\frac{d}{d\theta} \ln \theta^{N_1+\alpha} (1-\theta)^{N_0+\alpha} = 0.$$

$$\Rightarrow \frac{d}{d\theta} (N_1+\alpha) \ln\theta + (N_0+\alpha) \ln(1-\theta)$$

$$\Rightarrow \frac{(N_1+\alpha)}{\theta} - \frac{(N_0+\alpha)}{1-\theta} = 0$$

$$\Rightarrow (N_1+\alpha)(1-\theta) - N_0\theta - \alpha\theta = 0.$$

$$\Rightarrow N_1 + \alpha - N_1\theta - \alpha\theta = N_0\theta + \alpha\theta$$

$$\Rightarrow \theta = \frac{N_1+\alpha}{N_1+N_0+2\alpha}$$

3] $$p(c|x_i,\theta) = \frac{p(x_i|c,\theta)\cdot p(c)}{p(x_i)} \quad \cdots ①$$

From ①

$$p(c=1|x_i,\theta) = \frac{p(x_i|c=1,\theta)\, p(c=1)}{p(x_i)} \quad \cdots ②$$

$$p(c=2|x_i,\theta) = \frac{p(x_i|c=2,\theta)\, p(c=2)}{p(x_i)} \quad \cdots ③$$

~~Subtracti~~

Taking log on both sides of equation 2 and 3,

$$\log p(c=1|x_i,\theta) = \log \frac{p(x_i|c=1,\theta)\cdot p(c=1)}{p(x_i)} \quad \cdots ④$$

$$\log p(c=2|x_i,\theta) = \log \frac{p(x_i|c=2,\theta)\cdot p(c=2)}{p(x_i)} \quad \cdots ⑤$$

Subracting ⑤ from ④,

$$\log \frac{p(c=1|x_i,\theta)}{p(c=2|x_i,\theta)} = \log p(x_i|c=1,\theta) - \log p(x_i|c=2,\theta)$$

Since $p(c=1) = p(c=2) = 0.5$

$$\Rightarrow \log\left[\frac{p(c=1|x_i,\theta)}{p(c=2|x_i,\theta)}\right] = \phi(x_i)^T\beta_1 - \phi(x_i)^T\beta_2$$

... [From equation ② of assignment questions]

$$\log\left[\frac{p(c=1|x_i,\theta)}{\{p(c=2|x_i,\theta)\}}\right] = \phi(x_i)^T[\beta_1 - \beta_2]$$

b] For a word to have no effect on classification its corresponding terms in $\phi(x_i)^T(\beta_1 - \beta_2)$ must add up to zero.

$$\Rightarrow x_{1w}\beta_1 - x_{2w}\beta_2 = 0$$

$$\cancel{\Rightarrow x_{1w}\left[\log\frac{\theta_{1w}}{1-\theta_{1w}} + \sum\log(1-\theta_{1w})\right.}$$

$$x_{1w}\log\frac{\theta_{1w}}{1-\theta_{1w}} + \log(1-\theta_{1w}) - x_{2w}\log\left[\frac{\theta_{2w}}{1-\theta_{2w}}\right] - \log(1-\theta_{2w}) = 0.$$

Camlin Page
Date / /

~~also~~

$$x_{1w}\left[\log \frac{\theta_{1w}}{1-\theta_{1w}}\right] - x_{2w} \log \frac{\theta_{2w}}{1-\theta_{2w}} = \log(1-\theta_{2w}) - \log(1-\theta_{1w})$$

$x_{1w} = 1$ and $x_{2w} = 1$ since the word occurs in both documents ~~classes~~

$\Rightarrow \frac{\theta_{1w}(1-\theta_{1w})}{1-\theta_{1w}} = \frac{\theta_{2w}(1-\theta_{2w})}{1-\theta_{2w}}$ or equivalently $\beta_1 = \beta_2$

$\Rightarrow \theta_{1w} = \theta_{2w}$ i.e the probabilities of the word occuring in documents of both classes is equal.

c] From part (b), a word does not affect our beliefs about class label if its probability of occuring in both classes is same.

$\Rightarrow \hat{\theta}_{iw} = \hat{\theta}_{jw}$ for all $i,j \in C$

then the word is ignored by our classifier.

So even if $n_1 \neq n_2$

a word will be ignored if $\frac{1 + \sum_{i \in C=1} x_{iw}}{2 + n_1} = \frac{1 + \sum_{j \in C=2} x_{jw}}{2 + n_2}$ for some i, j

$$\Rightarrow \frac{\sum_{i \in C=1} x_{iw}}{n_1} = \frac{\sum_{i \in C=2} x_{iw}}{n_2}$$

therefore the ratio being same ensures that the word is ignored.

d] ~~0.25~~ The information gain from split on that word shoud ~~g~~ be minimum. One way is that it is equally probable across all ~~sets~~ classes.