Parameter Estimation Methods

① ⓐ probability of identical male twins : $\theta . p$

probability of identical female twins : $\theta(1-p)$

probability of fraternal male twins : $(1-\theta).q^2$

probability of fraternal female twins : $(1-\theta)(1-q)^2$

probability of fraternal opposite gender twins : $(1-\theta).2q(1-q)$

$$\text{Likelihood} = (\theta.p)^{m_i} . (\theta(1-p))^{f_i} . ((1-\theta)q^2)^{m_f} . ((1-\theta)(1-q)^2)^{f_f} . ((1-\theta)2q(1-q))^{b}$$

$$= \theta^{(m_i+f_i)} . (1-\theta)^{(m_f+f_f+b)} . p^{m_i} . (1-p)^{f_i} . q^{2m_f} . (1-q)^{2f_f} . (2q(1-q))^{b}$$

Scanned with CamScanner

$\log(likelihood) = (m_i + f_i).\log\theta + (m_f + f_f + b).\log(1-\theta) + m_i.\log p + f_i.\log(1-p)$

$2m_f.\log q + 2f_f.\log(1-q) + b.\log(2q(1-q))$

ⓑ $\frac{\partial \log(likelihood)}{\partial\theta} = \frac{m_i+f_i}{\theta} - \frac{m_f+f_f+b}{1-\theta} = 0 \longrightarrow \theta = \frac{m_i+f_i}{m_i+f_i+m_f+f_f+b}$

$\frac{\partial \log(likelihood)}{\partial p} = \frac{m_i}{p} - \frac{m_i+f_i}{1-p} = 0 \longrightarrow p = \frac{m_i}{m_i+f_i}$

$\frac{\partial \log(likelihood)}{\partial q} = \frac{2m_f+b}{q} - \frac{2f_f+b}{1-q} = 0 \longrightarrow q = \frac{2m_f+b}{2m_f+2f_f+2b}$

---

②

ⓐ likelihood $= p(x=4).p(x=2).p(x=7).p(x=9)$

$= (1-\theta)^3\theta\,(1-\theta)\theta\,(1-\theta)^6\theta\,(1-\theta)^8\theta = (1-\theta)^{18}.\theta^4$

$\log(likelihood) = 18\log(1-\theta) + 4\log\theta$

ⓑ $\frac{\partial \log(likelihood)}{\partial\theta} = 18\frac{-1}{1-\theta} + 4\frac{1}{\theta} = 0 \longrightarrow \theta = \frac{4}{22}$

if $\theta = \frac{4}{22}$ max? $\longrightarrow \frac{\partial^2 \log(likelihood)}{\partial\theta^2} = \frac{-18}{(1-\theta)^2} - \frac{4}{\theta^2} < 0$

Naive Bayes

①

ⓐ $Y_{prediction} = \underset{y}{\text{argmax}}\{\underbrace{p(y)}_{\text{prior}}.\underbrace{p(m|y).p(b|y).p(h|y)}_{\text{probability of feature given the class}}\}$

ⓑ $p(y=+1) = 4/6$ $\qquad p(m=1|y=+1) = 2/4$

$p(y=-1) = 2/6$ $\qquad p(m=1|y=-1) = 2/2$

$p(b=1|y=+1) = 2/4$ $\qquad p(h=1|y=+1) = 1/4$

$p(b=1|y=-1) = 1/2$ $\qquad p(h=1|y=-1) = 2/2$

ⓒ data : $(0, 0, 1)$

class $+1 = p(y=+1).p(m=0|y=+1).p(b=0|y=+1).p(h=1|y=+1)$

$= 4/6 . 2/4 . 2/4 . 1/4 = 1/24$ ⟵ our prediction

class $-1 = p(y=-1).p(m=0|y=-1).p(b=0|y=-1).p(h=1|y=-1)$

$= 0$

(d) $P(y=+1) = 5/8$ $\quad p(m=1|y=+1) = 3/6$ $\quad p(b=1|y=+1) = 3/6$

$P(y=-1) = 3/8$ $\quad p(m=1|y=-1) = 3/4$ $\quad p(b=1|y=-1) = 2/4$

$p(h=1|y=+1) = 2/6$

$p(h=1|y=-1) = 3/4$

class +1: $5/8 \cdot 3/6 \cdot 3/6 \cdot 2/6 = 5/96$ → our prediction $\quad$ class -1: $3/8 \cdot 1/4 \cdot 2/4 \cdot 3/4 = 9/256$

---

(2)

a)
p(y =spam).p(note =1|spam). p(to=0|spam).p(self=0|spam).p(become=0|spam).p(perfect=1|spam)
should be greater than
p(y =ham).p(note =1|ham). p(to=0|ham).p(self=0|ham).p(become=0|ham).p(perfect=1|ham)

p(ham) = 1- p(spam)

p(spam) > 0.2…

(b)

$$p(W = \text{Sir} \mid Y = \text{Spam}) = \frac{\#\ W = \text{Sir in Spams words}}{\#\ \text{words in spam}} = 1/13$$

$p(W = \text{watch} \mid Y = \text{ham}) = 1/9$

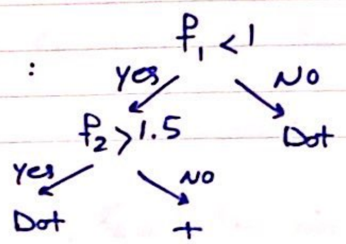
$p(W = \text{gauntlet} \mid Y = \text{ham}) = 0$ $\qquad p(Y = \text{ham}) = 2/3$

(c) $p(W=\text{sir} \mid Y=\text{spam}) = \dfrac{1+2}{13+2v}$ $\qquad p(Y=\text{ham}) = \dfrac{2+2}{3+2*2} = \dfrac{4}{7}$

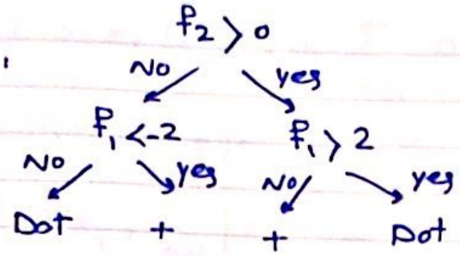
$p(W=\text{watch} \mid Y=\text{ham}) = \dfrac{1+2}{9+2v}$

Decision Trees

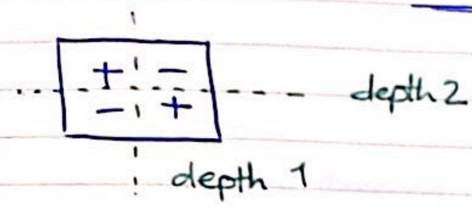
(1) Second plot :



Third plot : not possible with depth of 2 fourth plot



②



KNN

①

$|x_{12}| = 2$
$|x_{13}| = 2$
$|x_{14}| = 3$
$|x_{15}| = 3$
$\boxed{|x_{16}| = 1} \to ⊖$

prediction for x_1: ⊖

$|x_{21}| = 2$
$|x_{23}| = 5$
$|x_{24}| = 3$
$|x_{25}| = 3$
$\boxed{|x_{26}| = 1}$

prediction for x_2: ⊖

$\boxed{|x_{31}| = 2}$
$|x_{32}| = 5$
$|x_{34}| = 4$
$|x_{35}| = 4$
$|x_{36}| = 4$

prediction for x_3: ⊕

$|x_{41}| = 3$
$|x_{42}| = 3$
$|x_{43}| = 4$
$\boxed{|x_{45}| = 2}$
$|x_{46}| = 2$

prediction for x_4: ⊖

$|x_{51}| = 3$
$|x_{52}| = 3$
$|x_{53}| = 4$
$\boxed{|x_{54}| = 2}$
$|x_{56}| = 2$

prediction for x_5: ⊖

$\boxed{|x_{61}| = 1}$
$|x_{62}| = 1$
$|x_{63}| = 4$
$|x_{64}| = 2$
$|x_{65}| = 2$

prediction for x_6: ⊕

$$\text{accuracy} = \frac{3}{6} = \frac{1}{2} = 50\%$$

General ML Questions

1. overfitting
   both adding & removing features can help → remove noise / → add more informative data
   Collecting more data
   Increase k

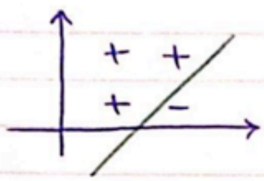
$\boxed{\text{Logistic Regression}}$:

①

② True

③ False : L_1 Regularization : $\max_{\theta}$ likelihood $(\theta, D) - \lambda \|\theta\|$

④



ⓐ Logistic Regression, We need a classifier not a regressor!

ⓑ yes, shown in the graph

if we change that example $\longrightarrow$ linear Logistic Regression cannot classify it correctly

ⓒ $P^+ = \frac{9}{9+1} = 0.9$

$R^+ = \frac{9}{9+9} = 0.5$