N. KAUSIK
CS21M037

Tutorial 6

2\. For the given data, we can fit a best fit line but we can see by plotting that error is high.

Hence we try fitting the data using a polynomial. If we take polynomial of degree 2, the data is fitted much better.

Hence we can consider variables $1, x, x^2$ and $y = \beta_0 + \beta_1 x + \beta_2 x^2$.
The model is still "Linear" in these transformed variable and thus basis expansion is done.

3\. a) Since we have $x_1 \dots x_n$ independent rand vars and all are Bernoulli,
for an x_i,

$$f(x_i ; p) = p^{x_i} (1-p)^{1-x_i}$$

and

Likelihood $L(p) = \prod_{i=1}^{n} f(x_i ; p)$ [as all x_i are independent]

$$= p^{x_1} (1-p)^{1-x_1} \cdots p^{x_n} (1-p)^{1-x_n}$$

$$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}$$

Log likelihood,

$$\log L(p) = (\log p) \sum_{i=1}^{n} x_i + (\log(1-p)) \left(n - \sum_{i=1}^{n} x_i\right)$$

diff wrt p and equate to 0,

$$\frac{\sum x_i}{p} - \frac{(n - \sum x_i)}{1-p} = 0$$

$\Rightarrow (1-p)\sum x_i = p(n - \sum x_i)$

$\Rightarrow \sum x_i - np = 0$

$\Rightarrow$ $$\boxed{\hat{p}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

4. We know, $P(\theta|S) = \dfrac{P(S|\theta)\,P(\theta)}{\int P(S|\theta)P(\theta)\,d\theta}$

Here, since we are using head or tail, $P(S|\theta)$ is from a Bernoulli dist.

Also since we did 20 trials and got Heads $\frac{1}{4} \times 20 = 5$ times. We got tails 15 times.

The prior dist is Beta (# heads, # tails)
$= \text{Beta}(5, 15)$.

$\therefore P(\theta|S) = (\text{const}) \cdot \theta^{y}(1-\theta)^{n-y}\,\theta^{4}(1-\theta)^{14}$

[$\int P(S|\theta)\,P(\theta)\,d\theta$ is const]

$= \theta^{y+4}(1-\theta)^{n-y+14} \Rightarrow \text{Beta}(y+5,\ n-y+15)$

$\therefore$ Likelihood $= \prod_{i=1}^{n} \theta^{x_i+4}(1-\theta)^{n-x_i+14}$

Log likelihood $= \sum_{i=1}^{n}\left((x_i+4)\log\theta + (n-x_i+14)\log(1-\theta)\right)$

diff wrt θ and equate to 0,

$\Rightarrow \dfrac{(\sum x_i + 4n)}{\theta} - \dfrac{(n^2+14n-\sum x_i)}{1-\theta} = 0$

$\Rightarrow \sum x_i + 4n - \theta(\sum x_i + 4n) = \theta(n^2+14n-\sum x_i)$
$\Rightarrow \sum x_i + 4n = \theta(n^2+14n-\sum x_i + \sum x_i + 4n)$

$\Rightarrow \boxed{\hat{\theta}_{MLE} = \dfrac{\sum_{i=1}^{n} x_i + 4n}{n^2+18n}}$