ML EXAM - 1

Name: Vyom Pathak

I.D.: 96703101

A1 Answer in Julia file

$$\theta = \begin{bmatrix} 1.14 \\ 0.71 \\ -0.71 \end{bmatrix}$$

```julia
X = [1 0 -2; 1 0 -1; 1 0 0; 1 1 0; 1 2 0];
y = [3;1;1;3;2];
theta = X\y;
display(theta);
```

A2

(a) Here $P(y)$ is denoted as the probability of seeing a particular class 'c'.

$\Rightarrow$ Thus $$P(y=i) = \frac{N_i}{N}$$

where, N_c = No. of documents belonging to class 'i', for $i = 1, \dots, K$

&

N = Total No. of documents.

(b) Consider the frequency count we have:

$x = [x_1, x_2, \dots, x_n]$.

⇒ When we combine this with the distribution of P with fraction of document belonging to each class 'i', and word 'j' at a word frequency of 'x':

$$P(i) \propto \pi_c \prod_{j=1}^{|V|} P(j/i)^{x_j}$$

⇒ In order to avoid underflow, we will use the sum of log's:

$$P(i) \propto \log\left(\pi_c \prod_{j=1}^{|V|} P(j|i)\right)$$

① ⇒ $P(i) = \log \pi_c + \sum_{j=1}^{|V|} x_j \log(P(j/i))$

© Here $\hat{y} = \max_c (w_c^T x + b_c)$

From the above

equation ①,
we get $b = \underline{\underline{\log(\pi)}}$

& $W_c = \log[\underline{\underline{P(c|i)}}]$

where $c = 1, \ldots, K$

A3 (a) Convex

(b) Convex

(c) Convex

(d) Not-Convex

(e) Convex

AS Here to incorporate the censored value, we introduce dummy variable as a placeholder. i.e. $y_{(i)}, i = p+1, \dots, n$. By introducing the dummy variables $(z_{(1)}, \dots, z_{(n-p)})$, we get the QP:

$$\text{minimize} \sum_{i=1}^{p} \left(y_{(i)} - w^T x_{(i)}\right)^2 + \sum_{i=p+1}^{n} \left(z_{(n-p)} - w^T x_{(i)}\right)^2$$

$$\text{subject to } z_{(i)} \geq D, \quad i = 1, \dots, n-p$$

where the variables are w and $z_{(i)}$, $i = 1, \dots, n-p$, which are optimized. Yes it is a convex problem

A6 $P(y|x) = \dfrac{\lambda^{y} e^{-\lambda}}{y!}$; $\lambda = e^{\phi^{T}\theta}$

Let there be 'n' datapoints such that $[y_1, y_2, \ldots, y_n]$ for respective $[x_1, x_2, \ldots, x_n]$.

$$\text{likelihood}(\theta) = \prod_{i=1}^{n} \frac{\lambda_i^{y_i} e^{-\lambda}}{y_i!}$$

$$\log(\text{likelihood}(\theta)) = \sum_{i=1}^{n} y_i \log \lambda_i - \lambda_i - \log y_i! \rightarrow ①$$

$\therefore$ MLE is defined as follows :

$$\text{maximize} \sum_{i=1}^{K} y_i \log \lambda_i - \lambda_i - \log y_i!$$

$\Rightarrow$ Taking the first derivative,

$$\frac{\partial f}{\partial \lambda} = \frac{\sum_{i=1}^{K} y_i}{\lambda} - n$$

$$\frac{\partial f}{\partial \lambda} = 0 \Rightarrow \lambda = \bar{y}$$

Taking second derivative:

$$\frac{d^2 f}{d\lambda} = -\frac{\sum_{i=1}^{K} y_i}{\lambda^2} \rightarrow (2)$$

Here, as (2) is negative, log-likelihood is concave, Hence, it is a convex optimization problem.

(A5) Lets consider the case when X is finite, from definition:

$$P(x;\theta) = a(\theta)\, e^{(\theta^T \phi(x))}$$

We have,

$$\log(p(x;\theta)) = \log(a(\theta)) + \theta^T \phi(x)$$

$$= -\log\left(\sum_{x \in X} e^{(\theta^T \phi(x))}\right) + \theta^T \phi(x)$$

$\Rightarrow$ Here 2^{nd} term in θ is affine, 1^{st} term is concave, as it's -ve-log-sum-exp function composed with affine function of theta $\theta^T \phi(x)$

$\Rightarrow$ With help of Hessian, we can show that $f(\theta) = \log(P(x;\theta))$ is concave [For more general case of a density f^{D}].

Defining $e(\theta, x) = \exp(\theta^T \phi(x))$

Second Partial derivative is:

$$\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} = \frac{\left(\left(\int_X \phi_i(x) e(\theta,x) dx\right)\left(\int_X \phi_j(x) e(\theta,x) dx\right) - \left(\int_X \phi_i(x)\phi_j(x) e(\theta,x) dx\right)\left(\int_X e(\theta,x) dx\right)\right)}{\left(\int_X e(\theta,x) dx\right)^2}$$

$\Rightarrow \phi x = (\phi_1(x), \cdots, \phi_n(x))$; for any vector $v \in \mathbb{R}^n$

$v^T(\nabla^2 f(\theta))\,v =$

$$\frac{\left(\int_x \left(\sum_i v_i \phi_i(x)\right) e(\theta,x)\,(dx)\right)^2 - \left(\int_x \left(\sum_i v_i \phi_i(x)\right)^2 e(\theta,x)\,dx\right)\left(\int_x e(\theta,x)\,dx\right)}{\left(\int_x e(\theta,x)\,dx\right)^2}$$

$\Rightarrow$ From Cauchy Schwarz inequality for function,

$g(x) = \left(\sum_i v_i \phi_i(x)\right) \sqrt{e(\theta,x)}$

$h(x) = \sqrt{e(\theta,x)}$

we get, $v^T(\nabla^2 f(\theta))\,v \leq 0$ for all v

Thus, as Hessian is negative semidefinite, so $\log(p(x;\theta))$ is concave.