Question 2:

(a) According to Normal Equation:

$$X^T(Y - h_\theta(X)) = 0$$

$\because$ $X_{m\times(n+1)}$, $Y_{(m\times 1)}$

$\therefore$

$$\begin{bmatrix} 1 & \cdots & 1 \\ x_1^1 & \ddots & \vdots \\ \vdots & & \\ x_1^m & \cdots & x_n^m \end{bmatrix}_{(n+1)\times m} \left(Y - h_\theta(X)\right) = 0$$

extract the first row: $\sum_{i=1}^{m} y^i = \sum_{i=1}^{m} h_\theta(x^i)$

$\because$ $h_\theta(x^i) = I\{y^i = 1\}$

$\therefore$ $\sum_{i=1}^{m} y^i = \sum_{i=1}^{m} I\{y^i = 1\} = \sum_{i=1}^{m} P(y=1 \mid x, \theta^*)$

(b) If a model is perfectly callibrated, then all we can say is that the probabilities output from the model match empirical observations. This is only about the possibilities of the outcomes, but not the outcomes themselves.

Conversely, for a perfectly accurate model, the probabilities output must match empirical results.

Question 3:

Suppose $\|\theta_{MAP}\|^2 > \|\theta_{ML}\|^2$

$\because$ prior $\theta \sim \mathcal{N}(0, \sigma^2 I)$

$\therefore$ $p(\theta_{MAP}) < p(\theta_{ML})$

$\therefore$ $p(\theta_{MAP}) \cdot \prod_{i=1}^{m} p(y^i | x^i; \theta_{MAP}) < p(\theta_{ML}) \cdot \prod_{i=1}^{m} p(y^i | x^i; \theta_{MAP}) < p(\theta_{ML}) \cdot \prod_{i=1}^{m} p(y^i | x^i; \theta_{ML})$

$\because \forall \theta, \prod_{i=1}^{m} p(y^i | x^i; \theta) < \prod_{i=1}^{m} p(y^i | x^i; \theta_{ML})$

$\Rightarrow$ False.

$\Rightarrow \|\theta_{MAP}\|^2 \le \|\theta_{ML}\|^2$

Question 4:

(a) $k(x,z) = k_1(x,z) + k_2(x,z)$

$\because$ $u^T k u = \underbrace{u^T k_1(x,z) u}_{>0} + \underbrace{u^T k_2(x,z) u}_{>0}$

$\therefore$ $u^T k u > 0 \Rightarrow k(x,z) > 0$

(b) $k(x,z) = k_1(x,z) - k_2(x,z)$

$\because$ $u^T k u = u^T k_1(x,z) u - u^T k_2(x,z) u$

$\therefore$ $u^T k u$ is not necessarily positive. $\Rightarrow$ k is not.

(c) $k(x,z) = a k_1(x,z)$

$\because$ $u^T k u = a u^T k_1(x,z) \cdot u$, a positive

$\therefore$ k is a kernel.

(d) $k(x,z) = -a k_1(x,z)$

contrary to question c, No.

(e) $k(x,z) = k_1(x,z)\, k_2(x,z)$

$$u^T k u = u^T k_1 k_2 u = u^T k_1 [uu^T][uu^T]^{-1} k_2 u$$

$$= u^T k_1 u \cdot u^T [uu^T]^{-1} k_2 u$$

$\because A^{-1} = [A^T A]^{-1} \cdot A^T$

$\therefore [uu^T]^{-1} = [[uu^T]^T [uu^T]]^{-1} \cdot [uu^T]^T$

$$= [[uu^T]^2]^{-1} \cdot [uu^T]^T$$

$\therefore u^T k u = u^T k_1 u \cdot u^T \cdot [[uu^T]^2]^{-1} \cdot [uu^T]^T k_2 u$

Let $C = [[uu^T]^2]^{-1}$, prove $C \geq 0$

$\because uu^T$ is symmetric $\Rightarrow [uu^T]^2$ is diagnalizable in an orthogonal basis:

$A = uu^T = Q \Lambda Q^{-1} \Rightarrow A^2 = [uu^T]^2 = Q \Lambda^2 Q^{-1} \Rightarrow$ semi definite positive.

$\Rightarrow$ $C = [(uu^T)^2]^{-1}$ is semi definite positive

$\Rightarrow$ $k \geq 0$

(f) $k(x, z) = f(x) f(z)$

No.

(g) $k(x, z) = k_3(\phi(x), \phi(z))$

$\because$ k_3 is Mercer

$\therefore$ k is Mercer

Question 5:

a. Let k be a Mercer kernel with mapping $\phi: \mathbb{R}^n \longmapsto E$

$K(x, y) \mapsto \langle \phi(x), \phi(y) \rangle$.

θ^i can be represented as $K(\theta^i, x^i)$

b. $h_{\theta^i}(x^{i+1}) = g(\theta^{i^T} \phi(x^{i+1})) = g(K(\theta^i, x^{i+1}))$

c. $K(\theta^{i+1}, x^{i+1}) := K(\theta^i, x^{i+1}) + \alpha y^{i+1} \cdot I\{y^{i+1} g(K(\theta^i, x^{i+1}))\} \cdot K(x^{i+1}, x^{i+1})$.

Question 6:

(a). find a way to compute NB predicted class.

$$r_i = e^{\log P(y_i=0|x_i) + \log(P(y=0)) - \log P(y_i=1|x_i) - \log(P(y=1))}$$

$$= \frac{P(y_i=0|x_i) \cdot P(y=0)}{P(y_i=1|x_i) \cdot P(y=1)}$$

$$p = \frac{1}{1+r_i} = \frac{P(y_i=1 \mid x_i)\,P(y=1)}{P(y_i=0 \mid x_i)\cdot P(y=0)+P(y_i=1 \mid x_i)\,P(y=1)}$$