Q1

(a) Let A = "patient is positive", B = "Test is positive"

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.95 * 0.0007}{0.95 * 0.0007 + 0.05 * 0.9993} = 0.0131$$

(b) $$P(A|B) = \frac{0.95 * 0.1}{0.95 * 0.1 + 0.05 * 0.9} = 0.679$$

(c) Let C = "patient is negative", D = "Test is negative"

Accuracy = $P(A) \cdot 95\% + P(C) \cdot 95\% = 95\%$

(d) $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.9 * 0.0007}{0.9 * 0.0007 + 0.001 \times 0.9993} = 0.3867$

if using P(patient is positive) = 0.1:

$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.9 * 0.1}{0.9 \times 0.1 + 0.001 * 0.9} = 0.99$

$\text{Accuracy} = 0.9 \times 0.0007 + 0.999 * 0.9993 = 0.9989$

(e) This test gives better true positive accuracy and overall accuracy.
when actual positive polulation has a low possibility.

(f) $\text{Accuracy} = 0.1 \times 0.0007 + 0.9999 * 0.9993 = 0.9993$

(g) $\text{Accuracy} = 0.1 * 0.05 + 0.9999 \times 0.95 = 0.955$

(h) No, although they're good at overall accuracy but often bad at true positives. Therefore, many positive patients cannot be detected.

Q2.

(a) Let random variable X denote observations of $x_i, i \in [1, N]$,

$$MSE = E[(X-a)^2] = E[X^2 - 2aX + a^2] = E[X^2] - 2E[aX] + E[a^2]$$

MSE is minimized when derivative with respect to a is 0:

$$-2E[X] + 2E[a] = 0, \text{ therefore } E[a] = E[X].$$

So, MSE is minimized when $a = E[X] = \frac{1}{N}\sum_{i=1}^{N} x_i$.

(b) $$\text{log-likelihood} = \log(P(X|a)) = \sum_{i=1}^{N} \log(N(a, \sigma^2))$$

$$= \sum_{i=1}^{N}\left(\log\left(\frac{1}{\sqrt{2\pi}\cdot\sigma}\right) - \frac{(x_i-a)^2}{2\sigma^2}\right)$$

Therefore, log likelihood is maximized when $\sum_{i=1}^{N}(x_i-a)^2$ is minimized.

(c) $$-\log B(y; \sigma(z)) = -\log\left[\sigma(z)^y (1-\sigma(z))^{1-y}\right]$$

$$= -y\log(\sigma(z)) - (1-y)\log(1-\sigma(z))$$

(d) $$\sum_{i=1}^{N} B(y_i; \sigma(b \odot x_i)) = \sum_{i=1}^{N} \sigma(b \odot x_i)^{y} (1-\sigma(b \odot x_i))^{1-y}$$

$$\sum_{i=1}^{N} -\log B(y_i; \sigma(b \odot x_i)) = \sum_{i=1}^{N} -y\log \sigma(b \odot x_i) - (1-y)\log(1-\sigma(b \odot x_i))$$

As the loss function is the same as negative log-likelihood for a Bernoulli random variable, maximize the likelihood of solution for a Bernoulli random variable is minimizing the loss function.