Q1)

(a) From marginalization and conditioning, we get

$y_A \sim \mathcal{N}(\mu_A, \Sigma_{AA})$

$y_B \sim \mathcal{N}(\mu_B, \Sigma_{BB})$

$$y_A | y_B \sim \mathcal{N}(\mu_A + \Sigma_{AB}\Sigma^{-1}_{BB}(y_B - \mu_B), \Sigma_{AA} - \Sigma_{AB}\Sigma^{-1}_{BB}\Sigma_{BA})$$

$$P(Y_i | Y_j = y_j, x_i, x_j) \sim \mathcal{N}(\mu_i + Cov(Y_i, Y_j)\, Var(Y_j)^{-1}(y_j - \mu_j), Var(Y_i) - Cov(Y_i, Y_j)\, Var(Y_j)^{-1} Cov(Y_j, Y_i))$$

$Y \sim \mathcal{N}(0, K)$, and we know

$$Cov(Y_i, Y_j)\, Var(Y_j)^{-1} = Corr(Y_i, Y_j) \cdot \frac{\sqrt{Var(Y_i)}}{\sqrt{Var(Y_j)}}$$

$$\therefore P(Y_i | Y_j = y_j, x_i, x_j) \sim \mathcal{N}\left(0 + 1 \cdot \sqrt{\frac{Var(Y_i)}{Var(Y_j)}}(y_j - 0), Var(Y_i) - Cov(Y_i, Y_j)\, Var(Y_j)^{-1} Cov(Y_j, Y_i)\right)$$

$Cov(Y_j, Y_i) = Cov(Y_i, Y_j)$

$$\therefore Cov(Y_i, Y_j)\, Var(Y_j)^{-1} Cov(Y_j, Y_i) = Corr(Y_i, Y_j)^2 \cdot Var(Y_i) = 1 \cdot Var(Y_i)$$

$\therefore$ From above, we have

$$P(Y_i | Y_j = y_j, x_i, x_j) \sim \mathcal{N}(0 + y_j, Var(Y_i) - Var(Y_i)) \sim \mathcal{N}(y_j, 0)$$

IIIly, $P(y_j | y_i = y_i, x_j, x_i) \sim \mathcal{N}(y_i, 0)$

<u>When $corr(Y_i, Y_j)$ approaches 0,</u>

$P(Y_j | Y_i = y_i, x_j, x_i)$ approaches $\mathcal{N}(0, Var(Y_j))$

and, $P(Y_i | Y_j = y_j, x_i, x_j)$ approaches $\mathcal{N}(0, Var(Y_i))$

Q1)

(b) (i) Given the kernel,

$$k_{PER}(x_i, x_j) = \exp\left(-2\sin^2\left(\pi/p\,|x_i - x_j|\right)\right)$$

$$\text{Corr}(Y_z, Y_{z+p}) = \frac{\text{Cov}(Y_z, Y_{z+p})}{\sqrt{\text{Var}(Y_z)\,\text{Var}(Y_{z+p})}} = \frac{k(Y_z, Y_{z+p})}{\sqrt{k(Y_z, Y_z)\,k(Y_{z+p}, Y_{z+p})}}$$

$$k_{PER}(x_i, x_i) = \exp\left(-2\sin^2\left(\frac{\pi}{p}|x_i - x_i|\right)\right) = \exp(0) = 1$$

$$\therefore \text{Corr}(Y_z, Y_{z+p}) = \text{Cov}(Y_z, Y_{z+p}) = k_{PER}(Y_z, Y_{z+p})$$

IIIly, $\text{Corr}(Y_z, Y_{z+2p}) = k_{PER}(Y_z, Y_{z+2p})$, and

$\text{Corr}(Y_{z+p}, Y_{z+2p}) = k_{PER}(Y_{z+p}, Y_{z+2p})$

$$\therefore \text{Corr}(Y_z, Y_{z+p}) = \exp\left(-2\sin^2\left(\pi/p\,|x_z - x_{z+p}|\right)\right)$$

$$= \exp\left(-2\sin^2\left(\pi/\cancel{p} \times \cancel{p}\,|x_i - x_{i+1}|\right)\right)$$

$$= \exp(0) = 1$$

IIIly, $\text{Corr}(Y_z, Y_{z+2p}) = \exp\left(-2\sin^2\left(\pi/p \times 2p\,|x_i - x_{i+1}|\right)\right)$

$$= \exp(0) = 1$$

and, $\text{Corr}(Y_z, Y_{z+2p}) = 1$

The above shows us that $Y_z, Y_{z+p}, Y_{z+2p}, \ldots$ are highly correlated,

and given that $Y_z = y_z$,

then $Y_{z+p} = Y_{z+2p} = Y_{z+3p} \ldots = y_z$

Q1)

(b)(ii) $k_{PER}(x_i, x_j) = \exp\left(-2\sin^2\left(\frac{\pi}{p}|x_i - x_j|\right)\right)$

Given $p = 4$

$x_1 = 1, \quad y_1 = 1$

$x_2 = 2, \quad y_2 = 2$

$x_3 = 3, \quad y_3 = 3$

$x_4 = 4, \quad y_4 = 2$

From b)(i), we see that $y_z = y_{z+p}$ when using the periodic kernel.

Since, $p = 4$,

$$y_i = y_{i+4} = y_{i+8} = y_{i+k(4)} \text{ where } k \in [1, \infty)$$

Thus,

$x_5 = 5 \Rightarrow y_5 = 1$

$x_6 = 6 \Rightarrow y_6 = 2$

$x_7 = 7 \Rightarrow y_7 = 3$

$x_8 = 8 \Rightarrow y_8 = 2$

$x_9 = 9 \Rightarrow y_9 = 1$

$x_{10} = 10 \Rightarrow y_{10} = 2$

$x_{11} = 11 \Rightarrow y_{11} = 3$

$x_{12} = 12 \Rightarrow y_{12} = 2$

Q1

(c) Basically from the question, if we want to show that $k_{sym}(X, Z)$ encodes reflective symmetry, we will want to show that the predictive mean satisfies $f(x) = f(-x)$ by showing that $Corr(Y_x, Y_{-x}) = 1$, which is equal to show $k_{sym}(x, x) = k_{sym}(-X, -X)$. We will show that it is axis-aligned reflective symmetry in the ~~flo~~ following:

$\because k_{sym}(X, Z) = k_b(X, Z) + k_b(-X, Z)$

$\therefore \begin{cases} k_{sym}(X, X) = k_b(X, X) + k_b(-X, X) \\ k_{sym}(-X, -X) = k_b(-X, -X) + k_b(X, -X) \end{cases}$

$\because$ for any X, Z, $k_b(X, X) = k_b(Z, Z) = S$, suppose $Z = -X$

$\therefore k_b(X, X) = k_b(-X, -X)$. $\because k(X, Z) = k(Z, X)$ for any kernel

$\therefore k_{sym}(X, X) = k_{sym}(-X, -X)$

$\because Cov(Y_i, Y_j) = k(X_i, X_j) \Rightarrow Cov(Y_x, Y_{-x}) = k_{sym}(X, -X)$

$$\therefore Corr(Y_x, Y_{-x}) = \frac{Cov(Y_x, Y_{-x})}{\sqrt{Var(Y_x) Var(Y_{-x})}} = \frac{k_{sym}(X, -X)}{\sqrt{k_{sym}(X, X)\, k_{sym}(-X, -X)}} = \frac{k_{sym}(X, -X)}{k_{sym}(X, X)}$$

$k_{sym}(X, -X) = k_b(X, -X) + k_b(-X, -X) = k_b(-X, X) + k_b(X, X)$

$k_{sym}(X, X) = k_b(X, X) + k_b(-X, X) \Rightarrow k_{sym}(X, -X) = k_{sym}(X, X)$

$\therefore Corr(Y_x, Y_{-x}) = 1$ $\because m(X) = f(X) = E(X)$, $k(X, X') = E[(X - m(X))(X' - m(X'))]$

$\therefore k_{sym}(X, Z)$ encodes reflective symmetry.

Question 2.

a) We start by taking the derivative of the loss function w.r.t T_L, and solve fo the min.

$$\frac{\partial L(s)}{\partial T_L} = 2\sum_{i\in L} w_i (y_i - T_L)(-1) = 0.$$

$$= \sum_{i\in L} w_i y_i - \sum_{i\in L} w_i T_L = 0$$

$$\boxed{\frac{1}{W_L}\sum w_i y_i = T_L}$$

$$\frac{\partial L(s)}{\partial T_R} = 2\sum w_i (y_i - T_R)(-1) = 0.$$

$$\sum w_i y_i - \sum w_i T_R = 0.$$

$$\boxed{\frac{1}{W_R}\sum w_i y_i = T_R}$$

There fore, choosing $T_R = \frac{1}{W_R}\sum_{i\in R} w_i y_i$ and $T_L = \frac{1}{W_L}\sum_{i\in L} w_i y_i$ minimizes the loss function.

b) we need three steps:

split the data: $N-1$ splits

For each split, compute loss for both sides for each point $N \leftarrow \sum w_i (y_i - T_L)^2 + \sum w_i (y_i - T_R)^2$

In total, we need $(N-1)N$ computations, we means we have $O(N^2)$ complexity

2. c)

$$L(s) = \sum_{i \in L} w_i (y_i - T_L)^2 + \sum_{i \in R} w_i (y_i - T_R)^2$$

$$= \sum_{i \in L} w_i (y_i^2 + T_L^2 - 2y_i T_L) + \sum_{i \in R} w_i (y_i^2 + T_R^2 - 2y_i T_R)$$

$$= \sum_{i \in L} w_i y_i^2 + \sum_{i \in L} w_i T_L^2 - \sum_{i \in L} 2w_i y_i T_L$$

$$+ \sum_{i \in R} w_i y_i^2 + \sum_{i \in R} w_i T_R^2 - \sum_{i \in R} 2w_i y_i T_R$$

$$= Q_L^{(k)} + \sum_{i \in L} w_i T_L^2 - 2P_L^{(k)} T_L$$

$$+ Q_R^{(k)} + \sum_{i \in R} w_i T_R^2 - 2P_R^{(k)} T_R$$

$$= \underline{Q_L^{(k)}} + \underline{W_L^{(k)} T_L^2} - \underline{2P_L^{(k)} T_L} + \underline{Q_R^{(k)}} + \underline{W_R^{(k)} T_R^2} - \underline{2P_L^{(k)} T_L}.$$

All terms are given, the sum of constant-time operations will result in constant time complexity

d) $x_1 \quad x_2 \quad x_3 \cdots x_N$. By moving the splitting point to the right by one, one point leaves the right partition to join the left, we can there for add one to the left and subtract one from the right.

1) $W_L^{k+1} = W_L^k + w_k \qquad W_R^{k+1} = W_R^k - w_k$.

2) $P_L^{k+1} = P_L^k + w_k y_k \qquad P_R^{k+1} = P_R^k - w_k y_k$.

3) $Q_L^{k+1} = Q_L^k + w_k y_k^2 \qquad Q_R^{k+1} = Q_R^k - w_k y_k^2$.

e) The complexity without the optimization is $O(N^2)$, with $O(N-1)$ for each split, and $O(N)$ for computing the loss function. Since the optization is just 4 additions and subtractions, complexity for this step is constant.

The new complexity is just $\boxed{O(N)}$

If sorting the vectors is counted, the complexity is $O(N) + O(N \log N)$.

Problem 3

1) By introducing a new bit to represent missing value, we allow for a new class to be split on.

2) We can impute missing values by borrowing strength from entries that do not have missing data. For example, if we have a feature that contains continuous numbers, we can use a regression algorithm with all the other data points to predict the values. In this case, we have $N-K$ values to be the training set for the regression where N is the total number of data points, and K is the number of points with missing value.