5.

From the problem desciption, we have the primal-dual sol:

$$(b^*, \tilde{w}^*, \alpha^*)$$

obtained by using $z_n^T = [1\ x_n^T]^T$ instead of x_n^T.

The $\tilde{w}^*$ is a $d+1$ dimensional vector:

$$\tilde{w}^* = [b^*, \tilde{w}_1^*, \cdots, \tilde{w}_d^*]$$

If we plug in this solution into primal problem P_1, we will get the minimized obejctive value:

$$\frac{1}{2}\tilde{\mathbf{w}}^{*T}\tilde{\mathbf{w}}^* + C\sum_{n=1}^{N}\xi_n$$

subject to:

$$y_n(\tilde{\mathbf{w}}^{*T} z_n) \geq 1 - \xi_n, \quad \forall n = 1, \cdots, N$$

We got the above inequality since we have:

$$\begin{aligned}
\tilde{\mathbf{w}}^{*T} z_n &= \begin{bmatrix} b^* & \tilde{w}_1^* & \cdots & \tilde{w}_d^* \end{bmatrix} \begin{bmatrix} 1 \\ x_{n1} \\ \vdots \\ x_{nd} \end{bmatrix} \\
&= \begin{bmatrix} \tilde{w}_1^* & \cdots & \tilde{w}_d^* \end{bmatrix} \begin{bmatrix} x_{n1} \\ \vdots \\ x_{nd} \end{bmatrix} + b^* \\
&= \mathbf{w}^{*T}\mathbf{x}_n + b^*
\end{aligned}$$

From the objective function of P1, we can see that:

$$\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N}\xi_n$$

is convex, because $\mathbf{w}^T\mathbf{w}$ is quadratic, and $C\sum_{n=1}^{N}\xi_n$ is linear (so also convex), and the sum of convex functions is convex.

And the first constraint:

$$y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1 - \xi_n, \quad \forall n = 1, \cdots, N$$

Consider the 2 cases,

1. $y_n = 1$:

$$\begin{aligned} \mathbf{w}^T\mathbf{x}_n + b &\geq 1 - \xi_n \\ \mathbf{w}^T\mathbf{x}_n &\geq 1 - \xi_n - b \end{aligned}$$

2. $y_n = -1$:

$$\begin{aligned} -\mathbf{w}^T\mathbf{x}_n - b &\geq 1 - \xi_n \\ \mathbf{w}^T\mathbf{x}_n &\leq -1 + \xi_n - b \end{aligned}$$

The 2 results both define a halfspace, and halfspaces are convex.

For the second constraint:

$$\xi_n \geq 0$$

This is also a halfspace, so this constraint is also convex.

Therefore, the primal problem P_1 is convex.

Since if any convex optimization problem with differentiable objective and constraints satisfies Slater's condition, then the KKT conditions are sufficient and necessary for optimality.

For prove that Slater's condition holds, we need to find a point $(\mathbf{w}^0, b^0, \xi^0)$ such that if:

$$\begin{aligned} g_n(\mathbf{w}, b, \xi) &= -y_n(\mathbf{w}^T\mathbf{x}_n + b) + 1 - \xi_n \leq 0, \quad \forall n = 1, \cdots, N \\ h_n(\xi) &= -\xi_n \leq 0, \quad \forall n = 1, \cdots, N \end{aligned}$$

$(\mathbf{w}^0, b^0, \xi^0)$ satisfies:

$$\begin{aligned} g_n(\mathbf{w}^0, b^0, \xi^0) &< 0 \\ h_n(\xi^0) &< 0 \end{aligned}$$

Let $(\mathbf{w}^0, b^0, \xi^0) = (\mathbf{0}, 0, 1.5)$, then:

$$\begin{aligned} g_n(\mathbf{0}, 0, 1.5) &= -y_n(0+0) + 1 - 1.5 = -0.5 < 0, \quad \forall n = 1, \cdots, N \\ h_n(1.5) &= -1.5 < 0 \quad \forall n = 1, \cdots, N \end{aligned}$$

Thus, Slater's condition holds.

Finally, to claim that $(b^*, \mathbf{w}^*, \alpha^*)$ is also a solution to P_1, we need to show that the KKT conditions are satisfied.

KKT conditions:

1. primal feasibility:

$$\begin{aligned} y_n(\mathbf{w}^{*T}\mathbf{x}_n + b^*) \geq 1 - \xi_n^*, \quad \forall n = 1, \cdots, N \\ \xi_n^* \geq 0, \quad \forall n = 1, \cdots, N \end{aligned}$$

By our derivation above, we know that:

$$y_n(\mathbf{w}^{*T}\mathbf{x}_n + b^*) \geq 1 - \xi_n^*, \quad \forall n = 1, \cdots, N$$

since $\tilde{\mathbf{w}}^{*T} z_n = \mathbf{w}^{*T}\mathbf{x}_n + b^*$.

Also, $\xi_n^* \geq 0$ is satisfied because it is the same as what Dr.Threshold get in his optimal solution.

2. dual feasibility:

Again, since α^* is the same as what Dr.Threshold get in his optimal solution, the following still holds:

$$0 \leq \alpha_n^* \leq C, \quad \forall n = 1, \cdots, N$$

3. complementary slackness:

From Dr.Threshold's solution, we know that:

$$\alpha_n^*(1 - \xi_n^* - y_n(\tilde{\mathbf{w}}^{*T}\mathbf{z}_n)) = 0, \quad \forall n = 1, \cdots, N$$

and:

- if $\alpha_n^* > 0$, then $1 - \xi_n^* - y_n(\tilde{\mathbf{w}}^{*T}\mathbf{z}_n) = 0$
- if $\alpha_n^* = 0$, then $1 - \xi_n^* - y_n(\tilde{\mathbf{w}}^{*T}\mathbf{z}_n) < 0$.

So converting the above equation, we can rewrite it as:

$$\alpha_n^*(1 - \xi_n^* - y_n(\mathbf{w}^{*T}\mathbf{x}_n + b^*)) = 0, \quad \forall n = 1, \cdots, N$$

Similar results can be obtained:

- if $\alpha_n^* > 0$, then $1 - \xi_n^* - y_n(\mathbf{w}^{*T}\mathbf{x}_n + b^*) = 0$
- if $\alpha_n^* = 0$, then $1 - \xi_n^* - y_n(\mathbf{w}^{*T}\mathbf{x}_n + b^*) < 0$.

Thus, KKT holds and $(b^*, \mathbf{w}^*, \alpha^*)$ is also an optimal solution to the original problem.