(a) Claim: The optimal vector $\mathbf{w}$ for the hard SVM problem will be a linear combination $a_1\mathbf{x}_1 + ... + a_N\mathbf{x}_N$ of the training data vectors $\mathbf{x}_1...\mathbf{x}_N$. (labels $y_1...y_N$).

(b) Proof: Suppose otherwise; then $\mathbf{w} = \mathbf{x} + \mathbf{o}$, where $\mathbf{x}$ is a linear combination of the training vectors and $\mathbf{o}$ is orthogonal to all the training vectors $\mathbf{x}_1...\mathbf{x}_N$, with $\mathbf{o} \neq \mathbf{0}$ (otherwise $\mathbf{w}$ would be a linear combination of the training vectors).[1] This will lead to a contradiction.

The hard SVM solution $\mathbf{w}, b$ must satisfy

$$\forall i, y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \tag{1}$$

Therefore, substituting $\mathbf{w} = \mathbf{x} + \mathbf{o}$ and using linearity,

$$\forall i, y_i(\mathbf{x}^T\mathbf{x}_i + \mathbf{o}^T\mathbf{x}_i + b) \geq 1$$

However, $\mathbf{o}$ is orthogonal to all training vectors $\mathbf{x}_i$, so,

$$\forall i, y_i(\mathbf{x}^T\mathbf{x}_i + 0 + b) \geq 1$$

So, $\mathbf{w} = \mathbf{x}, b = b$ also satisfies (1).

The other requirement for the hard SVM solution is that, subject to (1), $\mathbf{w}$ is the optimal solution to

$$min\ \frac{1}{2}\mathbf{w}^T\mathbf{w} \tag{2}$$

This is the contradiction; $\mathbf{w} = \mathbf{x} + \mathbf{o}$ does not satisfy (2) as $\mathbf{w} = \mathbf{x}$ is a better solution to (2) that still satisfies (1). I claim that

$$\frac{1}{2}\mathbf{x}^T\mathbf{x} < \frac{1}{2}(\mathbf{x} + \mathbf{o})^T(\mathbf{x} + \mathbf{o})$$

because [2]

$$(\mathbf{x} + \mathbf{o})^T(\mathbf{x} + \mathbf{o}) = \mathbf{x}^T\mathbf{x} + 2\mathbf{x}^T\mathbf{o} + \mathbf{o}^T\mathbf{o}$$
$$2\mathbf{x}^T\mathbf{o} = 0 \text{ for orthogonal } \mathbf{o}$$
$$\mathbf{o}^T\mathbf{o} > 0 \text{ for } \mathbf{o} \neq \mathbf{0}$$

[1] We can always find orthogonal $\mathbf{o}$. Find a basis A of the subspace spanned by the training vectors $\mathbf{x}_1...\mathbf{x}_N$ and a basis B of the subspace orthogonal to the training vectors, then express $\mathbf{w}$ as a linear combination of the bases A, B. The linear combination of A will be $\mathbf{x}$; the linear combination of B will be $\mathbf{o}$.

[2] $\mathbf{x}^T\mathbf{o} = a_1\mathbf{x}_1^T\mathbf{o} + ... + a_N\mathbf{x}_N^T\mathbf{o} = 0 + ... + 0 = 0$

Therefore $\mathbf{w}$ must be a linear combination of the training vectors $\mathbf{x}_1 ... \mathbf{x}_N$. (Recall that $\mathbf{x}$ was defined as a linear combination of the training vectors).