Problem 2.1. Given a training set $S = \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^m \subset (\mathbb{R}^d \bigotimes \{0,1\})^m$, show that there exists a polynomial p_s such that $p(\mathbf{x}) = 0$ if and only if $h_s(\mathbf{x}) = 1$.

Solution: Let $\mathbf{z_1}, \mathbf{z_2}, ..., \mathbf{z_k}$ be all of the training examples in S so that $\forall 1 \leq i \leq k$, $h(\mathbf{z_i}) = 1$. Then, we can define the polynomial $p_s : \mathbb{R}^d \to \{0,1\}$ in the following manner:

$$p_s(\mathbf{x}) = p_s(x_1, ..., x_d) = -\prod_{i=1}^{k}\sum_{j=1}^{d}(x_j - \mathbf{z_{ij}})^2$$

Now, it's clear that $\forall \mathbf{x} \in \mathbb{R}^d$, $p_s(\mathbf{x}) \leq 0$ since $\prod_{i=1}^k \sum_{j=1}^d (x_j - \mathbf{z_{ij}})^2 \geq 0$, no matter our choice of $\mathbf{x}$. Next, we must show that $p(\mathbf{x}) = 0$ if and only if $h_s(\mathbf{x}) = 1$. Pick $\mathbf{x} \in \mathbb{R}^d$ arbitrarily. If $p(\mathbf{x})$ then $\prod_{i=1}^k \sum_{j=1}^d (x_j - \mathbf{z_{ij}})^2 = \sum_{j=1}^d (x_j - \mathbf{z_{1j}})^2 ... \sum_{j=1}^d (x_j - \mathbf{z_{kj}})^2 = 0$. Now, fix a $p \in [k]$. Then, $\sum_{j=1}^d (x_j - \mathbf{z_{pj}})^2 = 0$ iff $x_j = \mathbf{z_{pj}}$, $\forall 1 \leq j \leq d$, as $(x_j - \mathbf{z_{pj}})^2 > 0$ if $x_j \neq \mathbf{z_{pj}}$. Hence, in order for $p_s(\mathbf{x}) = 0$, we must have that $\exists \alpha \in [k]$ so that $\mathbf{z}_\alpha = \mathbf{x}$. Now, if $h_s(\mathbf{x}) = 1$ then $\exists \beta \in [k]$ so that $\mathbf{z}_\beta = \mathbf{x}$ by the construction of $\mathbf{z_1}, \mathbf{z_2}, ..., \mathbf{z_k}$. Furthermore, $p_s(_\beta) = 0$ by construction. So, p_s satisfies our criteria. Thus, the hypothesis class of thresholded polynomials may lead to overfitting. $\square$

Problem 2.2.