Ex 2.1

Take $\hat{y}$ = Vector of length $K$ corresponding to the probability that observation $x$ belongs to each of the classes $k \in K$. Note the additional assumption here that $\hat{y}$ are probabilities $(0 \le \hat{y} \le 1)$ in addition to the information we are given in the question ($\hat{y}$'s sum to 1)

W.T.S. $\max_{i \in K} \hat{y}_i = \min_k \|t_k - \hat{y}\|$

$$\hat{y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K)$$

$$t_1 = (1, 0, \ldots, 0)$$
$$t_k = (0, \ldots, 0, 1, 0, \ldots, 0)$$

$k^{th}$ Position

Proof:

$$\min_k \|t_k - \hat{y}\|$$
$$= \min_k \sum_{i=1}^{K} ((t_k)_i - \hat{y}_i)^2$$

No dependance on $K$

$$= \min_k \sum_i \left( (t_k)_i^2 - 2\hat{y}_i (t_k)_i + \hat{y}_i^2 \right)$$

$$= \min_k \sum_i \left( -2\hat{y}_i (t_k)_i + (t_k)_i^2 \right)$$

$$= \min_k \sum_i \left( -2\hat{y}_i (t_k)_i + 1 \right) \quad \text{as} \quad \sum_i (t_k)_i = 1 \ \forall K$$

$$= \min_k \sum_i -\hat{y}_i (t_k)_i$$

$$= \max_k \sum_i \hat{y}_i (t_k)_i$$

$$= \max_{i \in K} \hat{y}_i$$