6.3 c)

Since the variances of the input data distribution in a) is much higher and there is an overlap between the classes, the plot wouldn't be so structured, which means that there would be wrong class assignments especially around the point where $w^T \phi(x)$ becomes zero. If a linear classifier would be applied, the data structure would be more difficult to approximate which would result in even more false classifications.

6.4 Kullback-Leibler Distance

$$
D_{KL}(p||q) \ = \ \sum_{x \in \chi} p(x) \, log \frac{p(x)}{q(x)} \quad = \quad \mathbf{E}_p \Big[log \frac{p(X)}{q(X)} \Big] \quad \Big| \text{X is a random variable} \tag{1}
$$

$$
= \quad \mathbf{E}_p \Big[- log \frac{q(X)}{p(X)} \Big] \quad \Big| \text{use Jensens Inequality} \tag{2}
$$

$$
\geq \quad -log \, \mathbf{E}_p \Big[\frac{q(X)}{p(X)} \Big] \tag{3}
$$

$$
= \quad -log \sum_{x \in \chi} p(x) \, \frac{q(x)}{p(x)} \tag{4}
$$

$$
= \quad -log \sum_{x \in \chi} q(x) \tag{5}
$$

$$
= \quad -log \, 1 \ = \ 0 \tag{6}
$$