First Non-linear Classification
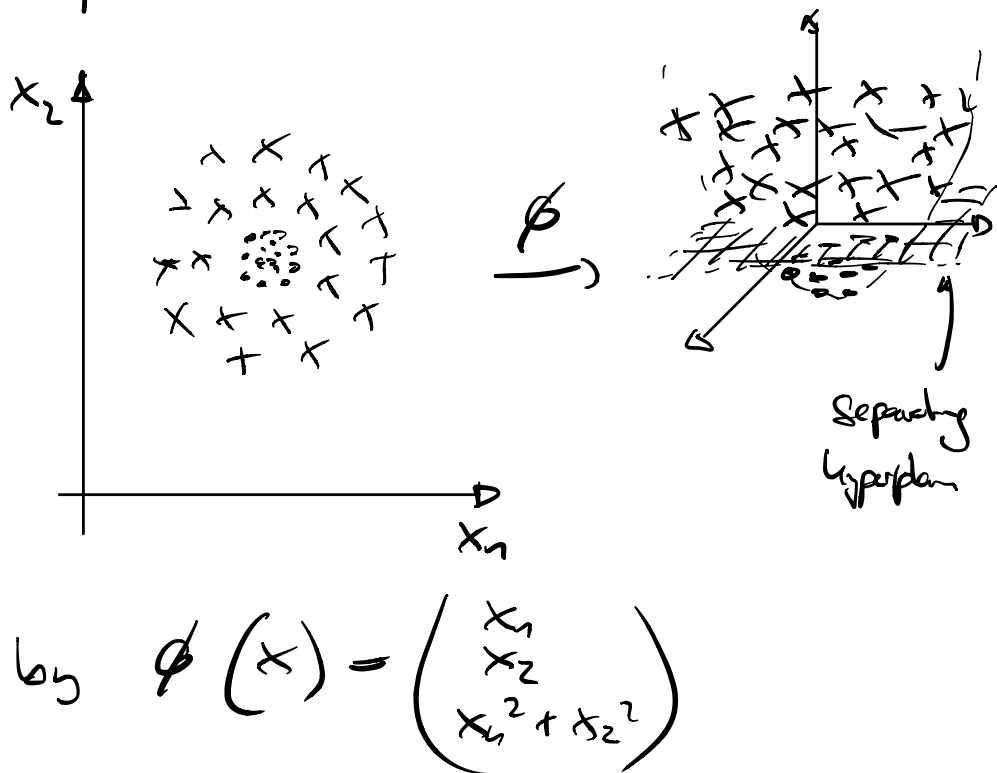
A first trick to extend our linear classifiers to non-linear ones D to map the training data points $x^{(i)}$ to a higher dimensional space where it becomes linearly separable

by $\phi(x) = \begin{pmatrix} x_1 \\ x_2 \\ x_1^2 + x_2^2 \end{pmatrix}$

Separating hyperplane

Hence, the activation become

$$\alpha(x) = w \cdot \begin{pmatrix} 1 \\ \underline{\phi}(x) \end{pmatrix}$$

for a suitable non-linear function

$$\underline{\phi} : \mathbb{R}^n \longrightarrow \mathbb{R}^h$$

and the loss function

$$L(w) = \frac{1}{M} \sum_{i=n}^{n} \ell(y^{(i)}, \alpha(x^{(i)}))$$

and the gradient

$$\frac{\partial L(w)}{\partial w} = \frac{1}{n} \sum_{i=n}^{M} \frac{d\ell(y^{(i)}, z)}{dz} \bigg|_{z = \alpha(x^{(i)})} \alpha'(x^{(i)})$$