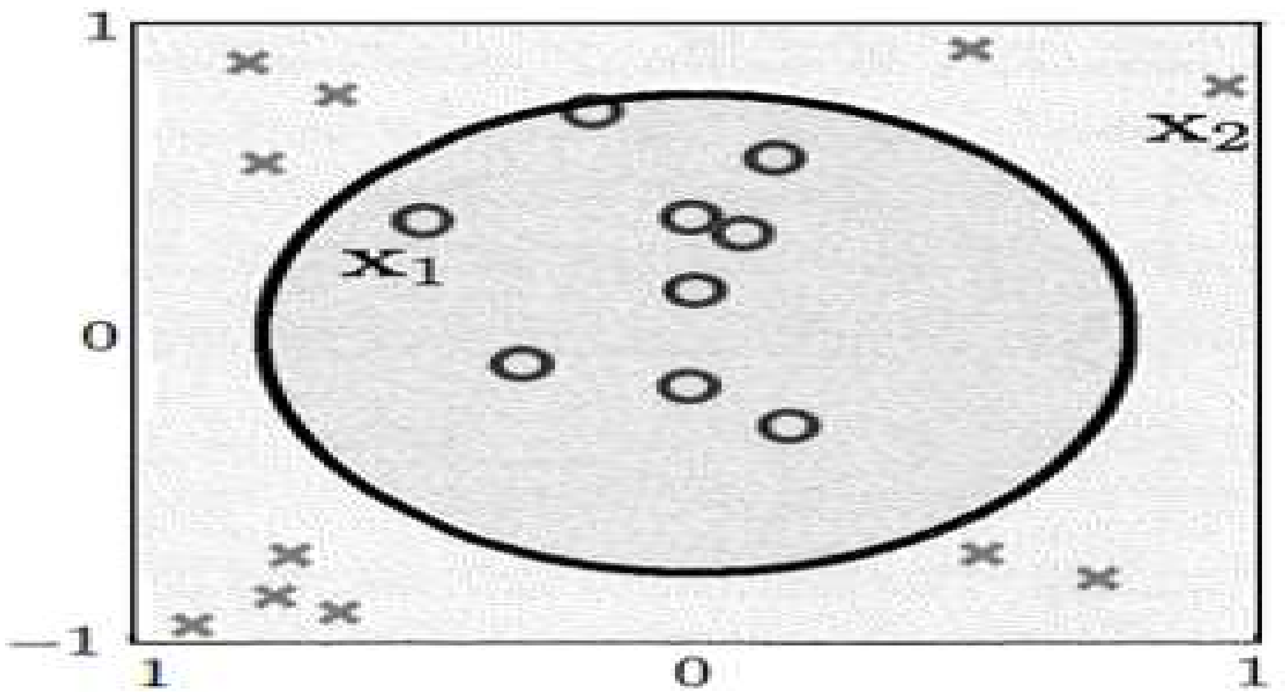
Classification
Transformations non-linéaires

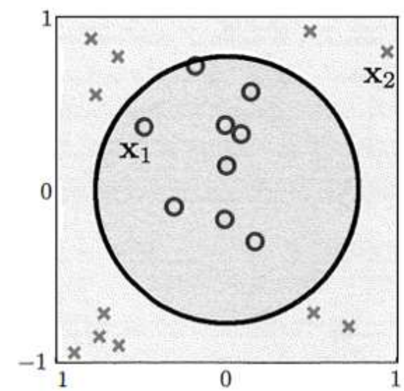
X transformer vers Z



Transformations non linéaires :Esapce X vers Espace Z

Considérons les données non linéairement séparables.

$$\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}) \rightarrow \boldsymbol{d_{cv}} = \mathbf{3}$$



- Le cercle présente l'équation suivante : $\boldsymbol{H} = \{\boldsymbol{c}, \boldsymbol{r} : \boldsymbol{c}, \boldsymbol{r} \in \mathbb{R}^{\mathbf{2}}\}$

$$C(x) = \boldsymbol{x_1^2} + \boldsymbol{x_2^2} - \mathbf{0.6} = \mathbf{0} \Longrightarrow \boldsymbol{c} = \mathbf{0}, \boldsymbol{r} = \sqrt{\mathbf{0,6}}$$

- Ainsi l'hypothèse non linéaire est :

$$\boldsymbol{h_{c,r}}(\boldsymbol{x}) = \boldsymbol{sign}(-\mathbf{0.6} + \boldsymbol{x_1^2} + \boldsymbol{x_2^2})$$

Cette hypothèse sera linéaire, après l'application d'une transformation non-linéaire sur les x_i.

- En particulier, considérons $z_0 = 1$, $z_1 = x_1^2$ et $z_2 = x_2^2$: $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}) \rightarrow \boldsymbol{z} = (\boldsymbol{z_0}, \boldsymbol{z_1}, \boldsymbol{z_2})$

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{sign}\left(\boldsymbol{w_0}.\underbrace{\boldsymbol{x_0}}_{\boldsymbol{z_0}} + \boldsymbol{w_1}.\underbrace{(\boldsymbol{x_1})^2}_{\boldsymbol{z_1}} + \boldsymbol{w_2}.\underbrace{(\boldsymbol{x_2})^2}_{\boldsymbol{z_2}}\right) \Longrightarrow \boldsymbol{h}(\boldsymbol{z}) = \boldsymbol{sign}(\widetilde{\boldsymbol{w}}_{\mathbf{0}}\boldsymbol{z_0} + \widetilde{\boldsymbol{w}}_{\mathbf{1}}\boldsymbol{z_1} + \widetilde{\boldsymbol{w}}_{\mathbf{2}}\boldsymbol{z_2})$$

Hyperplan: $\widetilde{\boldsymbol{w}}_{\mathbf{0}}\boldsymbol{z_0} + \widetilde{\boldsymbol{w}}_{\mathbf{1}}\boldsymbol{z_1} + \widetilde{\boldsymbol{w}}_{\mathbf{2}}\boldsymbol{z_2} \rightarrow \boldsymbol{d_{CV}} = \mathbf{4}$

- $x = (x_1, x_2) \rightarrow z = (z_0, z_1, z_2),\ w = (w_0, w_1, w_2)$
- $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{sign}\left(-\mathbf{0.6} + \boldsymbol{x}_{\mathbf{1}}^{\mathbf{2}} + \boldsymbol{x}_{\mathbf{2}}^{\mathbf{2}}\right) \boldsymbol{avec}\ (w_0 = -0{,}6\ , w_1 = 1, w_2 = 1)$
- $\boldsymbol{h}(\boldsymbol{z}) = \boldsymbol{sign}(\widetilde{\boldsymbol{w}}_{\mathbf{0}}\boldsymbol{z}_{\mathbf{0}} + \widetilde{\boldsymbol{w}}_{\mathbf{1}}\boldsymbol{z}_{\mathbf{1}} + \widetilde{\boldsymbol{w}}_{\mathbf{2}}\boldsymbol{z}_{\mathbf{2}})$

- $x \in X \subseteq \mathbb{R}^d \rightarrow x = (x_1, \cdots, x_d)$
- $z \in Z \subseteq \mathbb{R}^r\ \ avec\ r \geq d,\ \mathrm{z} = (z_1, \cdots, z_r)\ \ \ \phi_Q(X) = Z$

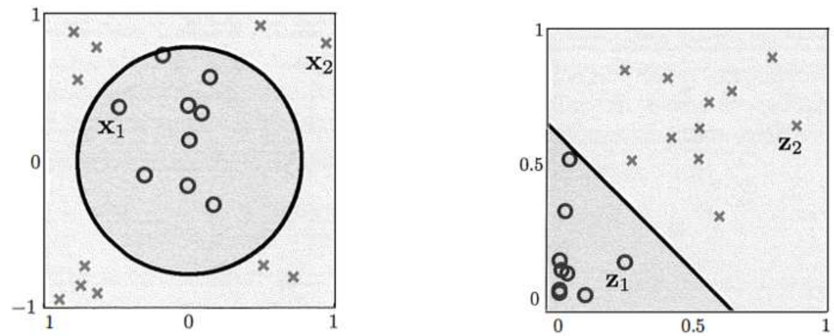
- $d = 2, Q = 2 \Longrightarrow x = (1, x_1, x_2), \phi_2(x) = \begin{pmatrix} 1 \\ x_1, x_2 \\ x_1^2, x_1\, x_{2,}\, x_2^2 \end{pmatrix} \Longrightarrow r = 6$

Transformations non linéaires : Espace Z

$$h(x) = sign(\widetilde{w}^T z)$$

$\widetilde{w}_i$: le poids dans l'espace des Z, et $\widetilde{d}$ est sa dimension.

Ainsi, on peut présenter les données en termes de z au lieu de x.



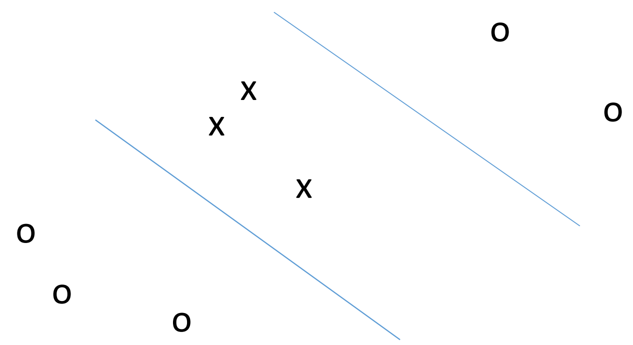
- L'espace Z contenant les vecteurs z, est nommé l'espace des caractéristiques.
- La transformé Φ qui lie X à Z est nommée « la transformée des caractéristiques »:

$$Z = \Phi(X)$$

PLA

- Si les données sont linéairement séparable alors
  - PLA converge
  - $d_{CV} = n + 1 < \infty \rightarrow$ on a l'apprentissage (PAC, APAC, CU)

- MLP



Transformations non linéaires : Espace Z

La forme générale de la transformée polynomiale des caractéristiques est:

$$\phi_Q(x) = \begin{pmatrix} 1, \\ x_1, \dots, x_d, \\ x_1^2, x_1 x_2, \dots, x_d^2, \\ \dots \\ x_1^Q, x_1^{Q-1} x_2, \dots, x_d^Q \end{pmatrix}$$

Q c'est l'ordre de la transformée.

Pour le cas de l'hypothèse cercle dans X: $x = (x_1, x_2) \rightarrow \mathrm{z} = \boldsymbol{\Phi}(\boldsymbol{x}) = (\mathbf{1}, \boldsymbol{x_1^2}, \boldsymbol{x_2^2})$

Toutes hypothèses non linéaires h (cercle) dans l'espace X peut être présentée par une hypothèse linéaire (hyperplan) dans l'espace Z :

$$h(x) = \tilde{h}(\boldsymbol{z}) = \tilde{h}(\boldsymbol{\Phi}(\boldsymbol{x}))$$

Transformations non linéaires : Espace Z

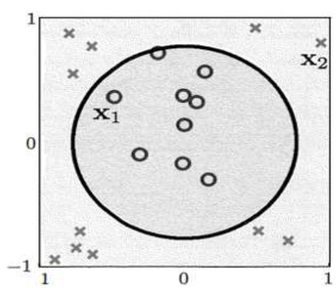
- L'ensemble des données transformées $(z_1, y_1), (z_2, y_2), \dots, (z_m, y_m)$ sont linéairement séparables dans Z.

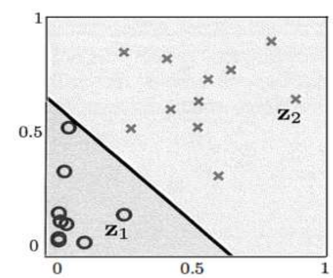
- Alors, on peut appliquer le PLA sur les données transformées pour obtenir $\tilde{w}_{PLA}$:

$$g(x) = sign(\tilde{w}_{PLA}^T z)$$

- L'erreur empirique dans l'espace X est la même que celle dans l'espace des caractéristiques Z, donc :

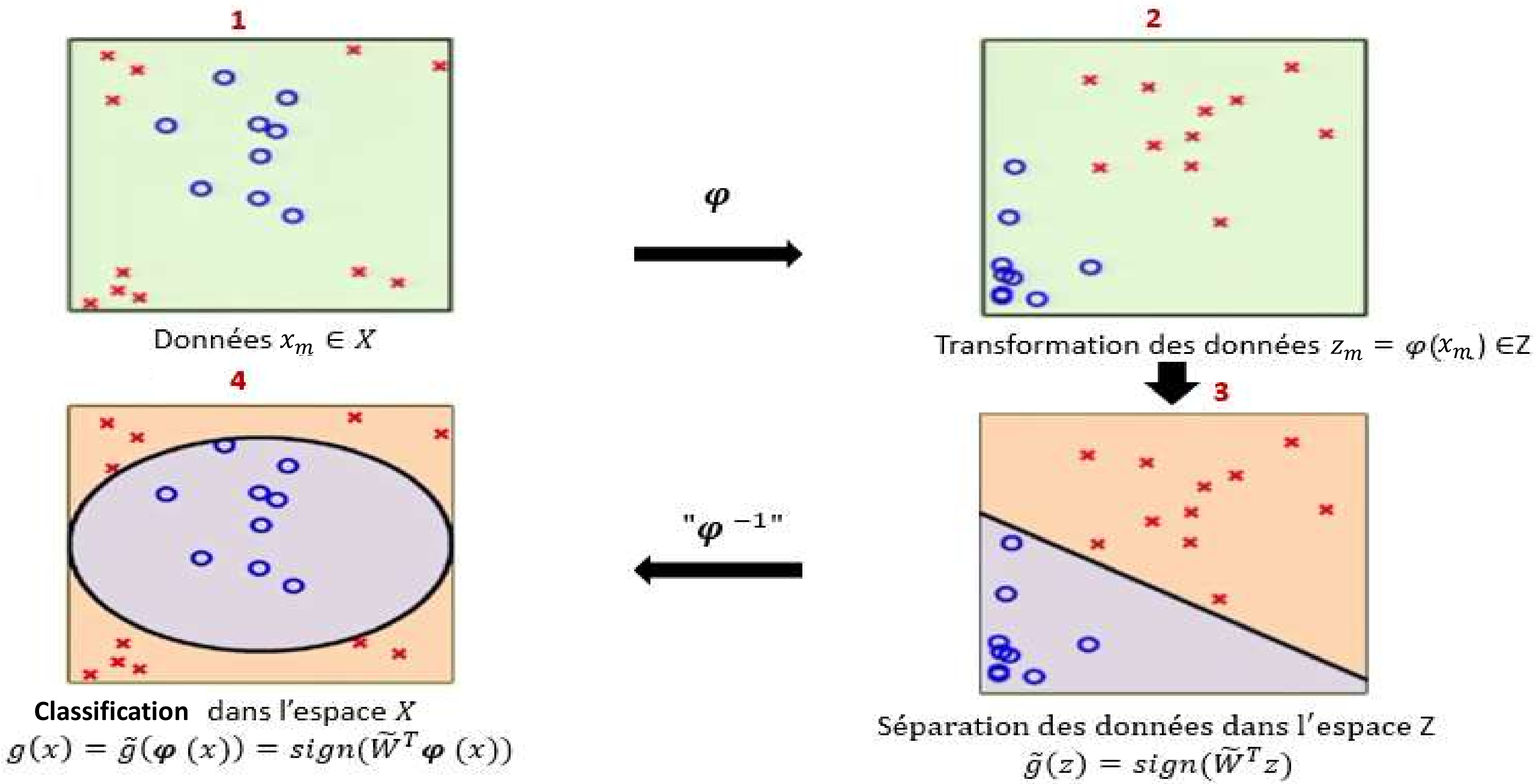
$$L_S(g) = 0$$





Transformations non linéaires : Espace Z

Le processus de la transformation des caractéristiques pour la classification linéaire:



- $\mathrm{z} = \boldsymbol{\Phi}'_2(\boldsymbol{x}) = \left(\mathbf{1}, \boldsymbol{x}_{\mathbf{1}}^{\mathbf{2}}, \boldsymbol{x}_{\mathbf{2}}^{\mathbf{2}}\right) \rightarrow \boldsymbol{d}_{\boldsymbol{CV}} = \mathbf{3}$
- $\mathrm{z} = \boldsymbol{\Phi}_2(\boldsymbol{x}) = (\mathbf{1}, \boldsymbol{x}_{\mathbf{1}}^{\mathbf{2}}, \boldsymbol{x}_{\mathbf{1}}\boldsymbol{x}_{\mathbf{2}}, \boldsymbol{x}_{\mathbf{2}}^{\mathbf{2}}) \rightarrow \boldsymbol{d}_{\boldsymbol{CV}} = \mathbf{4}$
- **Si on a overfitting reduire le degree de polynome**

- $x \in X \subseteq \mathbb{R}^d \rightarrow x = (x_1, \cdots, x_d)$
- $z \in Z \subseteq \mathbb{R}^r \ \ avec\ r = f(d)$

Transformations non linéaires : Espace Z

Revenons à l'exemple dont: $\mathrm{z} = \boldsymbol{\Phi}(\boldsymbol{x}) = \left(\mathbf{1}, \boldsymbol{x}_1^2, \boldsymbol{x}_2^2\right) = (\mathbf{1}, \boldsymbol{x}_1^2, \boldsymbol{x}_1\boldsymbol{x}_2, \boldsymbol{x}_2^2)$

On a : $Z = \{1\} \times \mathbb{R}^2$

Puisque H_Φ est l'ensemble des hypothèses d'un perceptron dans Z, donc:

$$\boldsymbol{d_{VC}(H_\Phi) \leq 3} \textbf{ et non pas } \boldsymbol{d_{VC}(H_\Phi) = 3}$$

?Car il existe des points $\boldsymbol{x} \in \boldsymbol{X}$ qui n'ont pas de transformés correctes dans $\boldsymbol{Z}$.

Remarque:

La transformée des caractéristiques doit être choisie avec précaution:

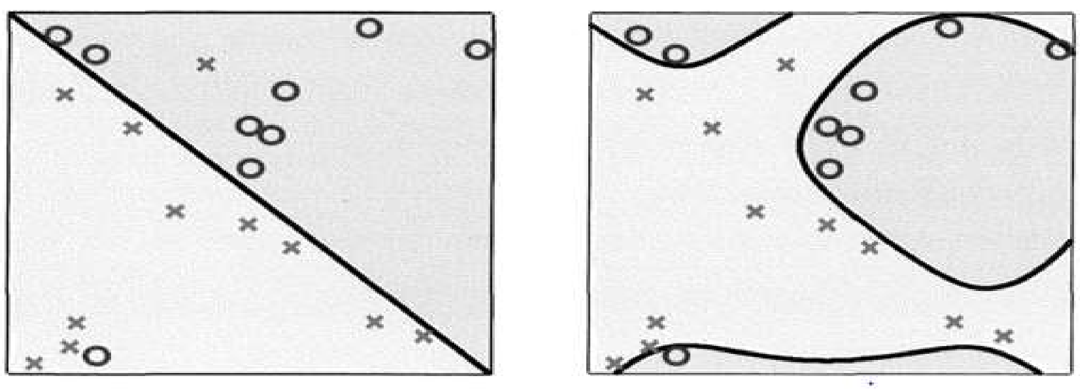
- Il faut choisir Φ avant de voir les données ou d'exécuter l'algorithme, afin de ne pas tomber dans le overfitting.
- Il ne faut pas insister sur la séparation linéaire et par la suite utiliser une **hypothèse très complexe**.

Transformations non linéaires : Espace Z

- Il n'existe aucune ligne qui peut séparer parfaitement les données d'entrainement, ni courbe quadratique, ni courbe polynomiale d'ordre trois.
- Pour cela, il faut utiliser une transformé polynomiale d'ordre $Q = 4$:

$$\Phi_4(x) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3, x_1^4, x_1^3x_2, x_1^2x_2^2, x_1x_2^3, x_2^4)$$

dans ce cas $\tilde{d} = 14$

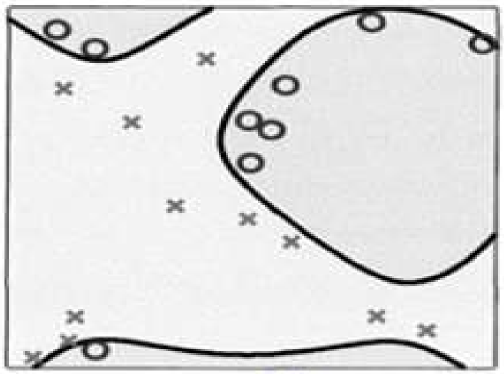


(a) Hypothèse linéaire. (b) hypothèse polynomiale d'ordre quatre.

Transformations non linéaires : Espace Z

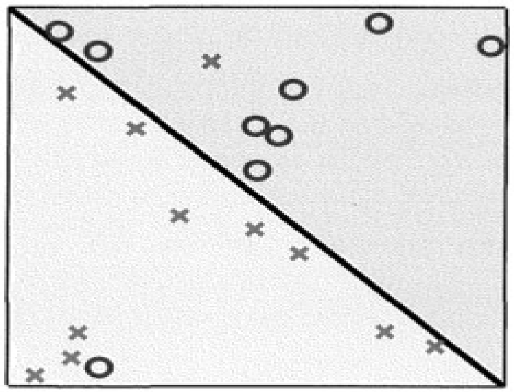
- Cette figure montre qu'on a surestimé les données:
  - La capacité de généralisation(erreur ↑) ↓
  - La capacité d'approximation(erreur ↓) ↑

$$L_S = 0$$



- La meilleure solution est d'ignorer les deux points mal classés:
  - La capacité de généralisation ↑
  - La capacité d'approximation ↓

$$L_S \neq 0$$



Remarque:

Il faut tolérer l'erreur empirique tout en choisissant une hypothèse très simple.

Transformations non linéaires : Limites

- L'utilisation d'une transformé polynomiale d'ordre très grand Q, nous donne beaucoup de flexibilité en termes de la forme des décisions dans X.

- Mais, il y a un prix à payé. Ce prix est celui de:

  - Mauvaise généralisation.

  - Complexité de calcul.

Transformations non linéaires : Complexité de calcul

- Le calcul est un grand problème.

- La transformé Φ_Q transforme:

Un vecteur x à deux dimensions $\boldsymbol{d = 2}$ en un vecteur à dimensions:

$$\tilde{\boldsymbol{d}} = \frac{\boldsymbol{Q(Q+3)}}{\boldsymbol{2}} \textbf{ ?}$$

- Ce qui augmente la complexité de la mémoire de calcul.

- Les choses peuvent s'aggraver si la dimension de x est très grande.

Apprentissage CU=APAC=PAC

- $|L_D(h) - L_S(h)| \approx 0$
- $L_D(h) \approx 0$

Transformations non linéaires : Généralisation

- Si Φ_Q est la transformée d'un espace d'entrée à deux dimensions, il existera $\tilde{d} = \frac{Q(Q+3)}{2}$ dimensions dans Z:

$$\boldsymbol{d_{VC}(H_Q)} \textbf{ sera proche de } \tilde{\boldsymbol{d}} + \boldsymbol{1} = \frac{\boldsymbol{Q(Q+3)}}{\boldsymbol{2}} + \boldsymbol{1}$$

- Cela veut dire que le second terme de la limite de généralisation peut augmenter d'une manière significatif:

$$|L_D(h) - L_S(h)| \leq \frac{4 + \sqrt{\log\left(\Pi_H(2m)\right)}}{\delta\sqrt{2m}}$$

Exemple:

si on utilise $\Phi = \Phi_{50}$, la dimension $d_{VC}(H_Q)$ sera approximativement égale à:

$$\tilde{d} + 1 = \frac{50(50+3)}{2} + 1 = 1326$$

Transformations non linéaires : Généralisation

Selon le théorème fondamental de l'apprentissage, on a:

$$C_1 \frac{d_{VC} + \log(\frac{1}{\delta})}{\varepsilon^2} \leq m_H^{APA} \quad (\varepsilon, \delta) \leq C_2 \frac{d_{VC} + \log(\frac{1}{\delta})}{\varepsilon^2}$$

Puisque $d_{VC}(H_Q)$ est très grande, donc on aura besoin des milliers de données par rapport au cas où on n'utilise pas de transformation.

Lors du choix de la dimension de la transformé des caractéristiques, on ne peut pas éviter le compromis approximation/généralisation:

- **Approximation :**

  Un $\tilde{d}$ très grand $(\mathbf{Q} \uparrow) \rightarrow (L_S) \downarrow$ et $(d_{VC}) \uparrow$

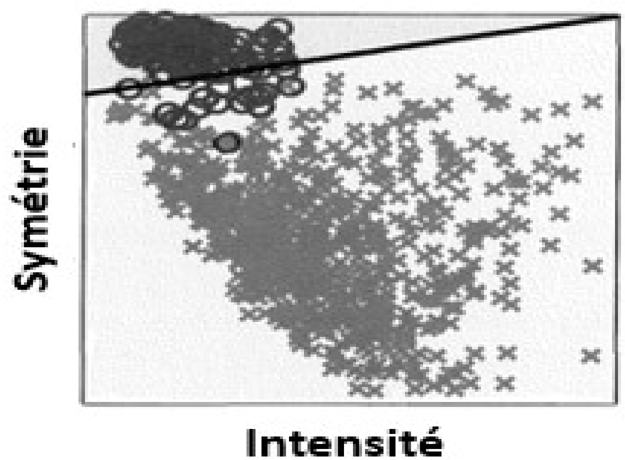
- **Généralisation :**

  Un $\tilde{d}$ très petit $(\mathbf{Q} \downarrow) \rightarrow (L_S) \uparrow$ et $(d_{VC}) \downarrow$?

Transformations non linéaires : Exemple

Exemple : Reconnaissance des chiffres

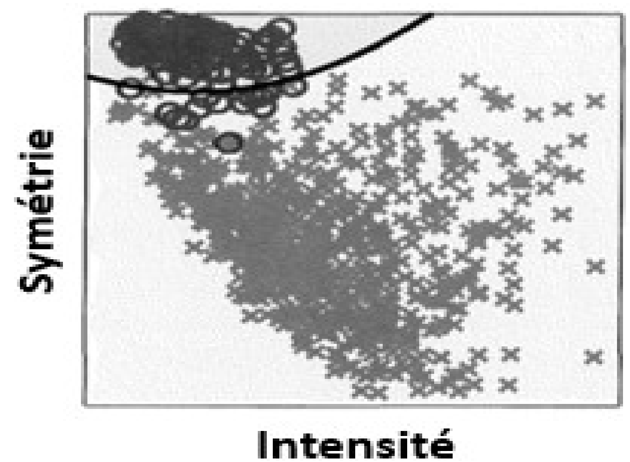
Une ligne sépare difficilement le chiffre 1 des autres chiffres, mais une courbe peut faire mieux.



Modèle linéaire

$L_S = 2.13\%$

$L_D = 2.38\%$



Modèle polynomial d'ordre 3

$L_S = 1.75\%$

$L_D = 1.87\%$