Exercice 1:

Mq que l'equation de la regression linéaire (MAP) de la suivante à l'aide de la représentation duale

$$J(\vec{w}) = \frac{1}{2}\sum_{n=1}^{N}(\vec{w}^T\phi(x_n)-t_n)^2 + \frac{\lambda}{2}\vec{w}^T.\vec{w} \quad (1)$$

est donnée suivant l'equation: $y(x) = K(x)^T(K+\lambda I)^{-1}t = a^T\phi(x)$

Soit $y(\vec{w}) = \frac{1}{2}\sum_{n=1}^{N}(\vec{w}^T\phi(x_n)-t_n)^2 + \frac{\lambda}{2}\|\vec{w}\|^2$ (la solution à posteriori)

En forçant le gradient à 0.

$$\frac{\partial J(\vec{w})}{\partial w} = 0 \iff \frac{2}{2}\sum_{n=1}^{N}(\vec{w}^T\phi(x_n)-t_n)\phi(\vec{x_n}) + \lambda\vec{w} = 0$$

$$\to \sum_{n=1}^{N}(\vec{w}^T\phi(x_n)-t_n)\phi(x_n) = -\lambda\vec{w}$$

$$\vec{w} = -\frac{1}{\lambda}\sum_{n=1}^{N}(\vec{w}^T\phi(x_n)-t_n)\phi(x_n) = -\sum_{n=1}^{N}\left(\frac{\vec{w}^T\phi(x_n)-t_n}{\lambda}\right)\phi(x_n)$$

on pose $a_n = -\frac{\vec{w}^T\phi(x_n)-t_n}{\lambda} \quad \forall n \in \mathbb{N}$

$$\to \vec{w} = \sum_{n=1}^{N}a_n\phi(x_n) \Rightarrow \boxed{\vec{w} = \phi^T\vec{a}} \quad (2)$$ tq $\phi = \begin{bmatrix}\phi(x_1)\\ \vdots \\ \phi(x_N)\end{bmatrix}$ et $\vec{a} = \begin{bmatrix}a_1\\ \vdots \\ a_N\end{bmatrix}$

De (1) et (2): $J(\vec{a}) = \frac{1}{2}[a^T\phi\phi^T\phi\phi^T\vec{a} - 2\vec{a}^T\phi\phi^T\vec{t} + \vec{t}^T\vec{t}] + [\frac{\lambda}{2}\vec{a}^T\phi\phi^T\vec{a}]$ (voir page suivante pour plus de details)

En introduisant la Matrice de gram tq $K = \phi\phi^T$ et $K(\vec{x_n},\vec{x_m}) = \phi(\vec{x_n})^T\phi(\vec{x_m})$

$$K = \begin{pmatrix} K(\vec{x_1},\vec{x_1}) & \dots & K(\vec{x_1},\vec{x_N}) \\ K(\vec{x_N},\vec{x_1}) & & K(\vec{x_N},\vec{x_N})\end{pmatrix}$$

donc $J(\vec{a}) = \frac{1}{2}[\vec{a}^TKK\vec{a}] - \vec{a}^TK\vec{t} + \frac{1}{2}\vec{t}^T\vec{t} + \frac{\lambda}{2}\vec{a}^TK\vec{a}$

En forçant le gradient à 0 On aura: $\vec{\nabla}J(\vec{a}) = 0 \iff \frac{2}{2}K.K.\vec{a} + \lambda.\frac{2}{2}.K\vec{a} = K\vec{t}$

$$\iff K\vec{a} + \lambda\vec{a} = \vec{t}$$

$$\iff \vec{a} = (K+\lambda I_n)^{-1}\vec{t}.$$

Pour la prédiction d'une entrée $\vec{x}$ donné On applique les fcts de bases à chaque donnée.

tq $\vec{x_n} \to \vec{\phi}(x_n) \quad \forall n$

$$y_w(\vec{x}) = \vec{w}^T\phi(\vec{x}) = \vec{a}^T\phi\phi(\vec{x}) = [(K+\lambda I_n)^{-1}\vec{t}]^T\phi\phi(\vec{x})$$

$$= [[(K+\lambda I_n)^{-1}\vec{t}]^T.\phi\phi(\vec{x})]^T \in \mathbb{R} \quad (y_w(\vec{x}) = y_w^T(\vec{x}))$$

$$= [(K+\lambda I_n)^{-1}\vec{t}]^T(\phi\phi(\vec{x}))^T$$

Or $(\phi.\vec{\phi}(\vec{x}))^T = [\phi(\vec{x})]^T.\phi$ et $\phi^T = (\phi(x_1), \phi(x_2) \dots, \phi(x_N))$

$$= [\phi(\vec{x})]^T(\phi(x_1), \dots, \phi(x_N))$$

$$= (\phi(x_1), \dots, \phi(x_N))[\phi(\vec{x})]^T$$

$$= (\phi(x_1)\phi^T(\vec{x}), \dots, \phi(x_N)\phi^T(\vec{x}))$$

$$= \vec{K}(\vec{x})^T$$

$$\to \boxed{y_{\vec{w}}(\vec{x}) = K^T(\vec{x})(K+\lambda I_n)^{-1}\vec{t}}$$ tq $K(\vec{x}) = \begin{pmatrix}K(\vec{x_1},\vec{x})\\ K(\vec{x_N},\vec{x})\end{pmatrix}$

De ① et ② $\begin{cases} J(\vec{w}) = \frac{1}{2}\sum_{w=1}^{N} (\vec{w}^T \phi(x_n) - t_n)^2 + \frac{\lambda}{2}\vec{w}^T\vec{w} \\ \vec{w} = \phi^T \vec{a} \end{cases}$

$$\Leftrightarrow J(\vec{w}) = \frac{1}{2}\sum_{n=1}^{N} (a^T\phi\,\phi(x_n) - t_n)^2 + \frac{\lambda}{2} a^T\phi\phi^T a$$

$$= \frac{1}{2}\sum_{n=1}^{W} \left((a^T\phi.\phi(x_n))^2 - 2a^T\phi.\phi(x_n)t_n + t_n^2 \right) + \frac{\lambda}{2} a^T\phi\phi^T a$$

$$= \frac{1}{2}\sum_{n=1}^{N} a^T\phi.\phi(x_n) \times a^T\phi.\phi(x_n) - \sum_{n=1}^{N} a^T\phi.\phi(x_n)t_n + \frac{1}{2}\sum_{n=1}^{N} t_n.t_n$$
$$+ \frac{\lambda}{2} a^T\phi\phi^T a$$

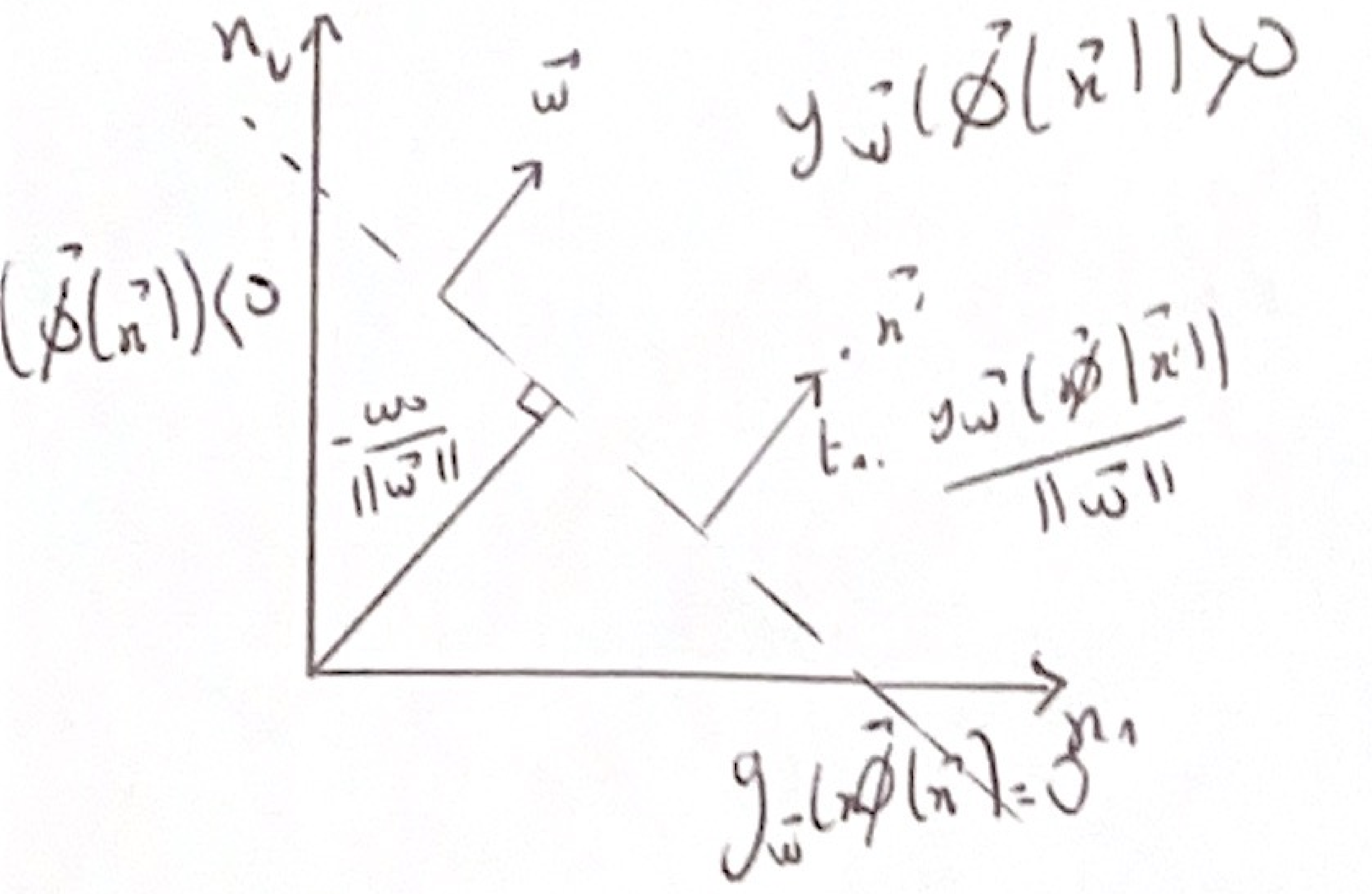
$$= \frac{1}{2} a^T\phi\phi^T\phi\phi^T a - a^T\phi\phi^T t + \frac{1}{2}t^T t + \frac{\lambda}{2} a^T\phi\phi^T a$$

On pose $K = \phi\phi^T$

$$\Rightarrow J(\vec{a}) = \frac{1}{2} a^T K K a - a^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T K a.$$

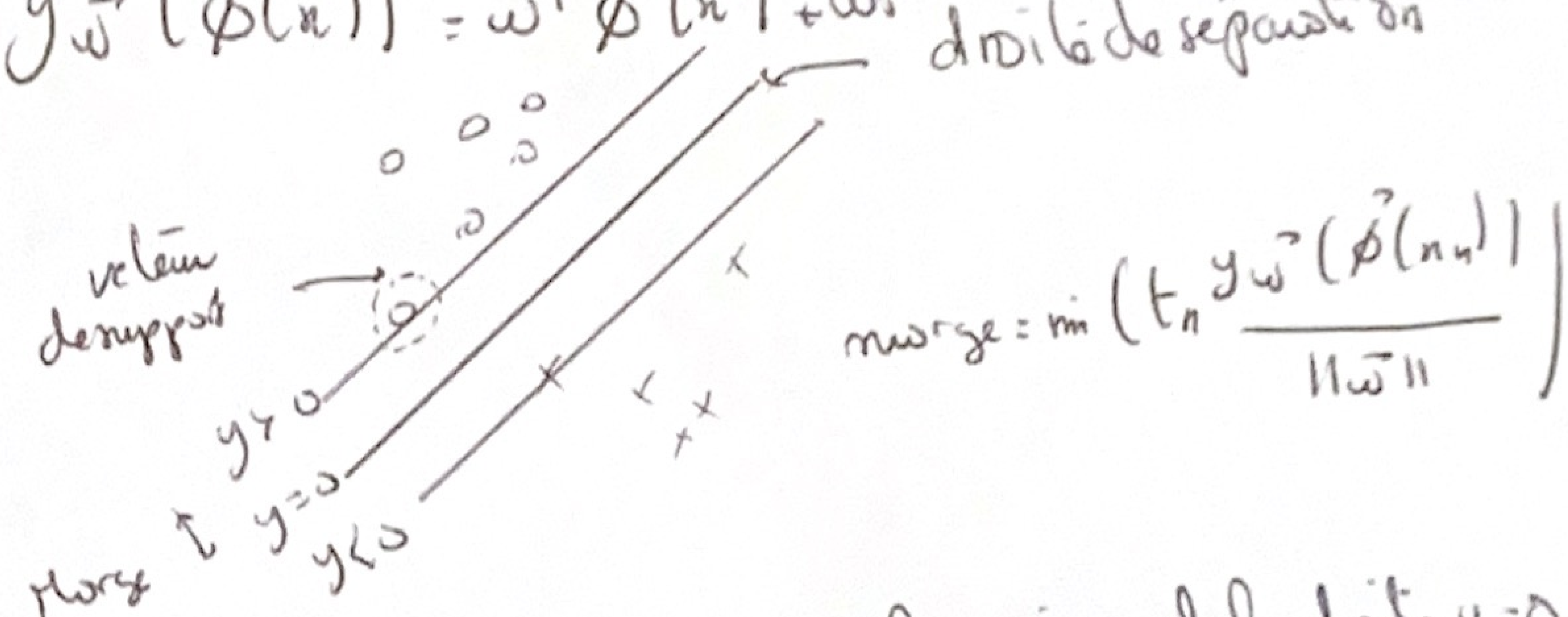
Exo 2:

Avant d'expliciter les vecteurs de support il est primordial d'aborder la notion de marge. En effet, la marge est la plus petite distance entre la surface de séparation et les données d'entraînement.



En considérant la fct de base:

$$y_{\vec{w}}(\vec{\phi}(\vec{x})) = \vec{w}^T \vec{\phi}(\vec{x}) + w_0$$



Les vecteurs support alors sont les points qui tombent sur la marge qui leur sépare de la droite $y = 0$.

En utilisant une approche avec la méthode des noyaux on calcule la prédiction

$$(*)\quad y(\vec{x}) = \sum_{n=1}^{N} a_n t_n k(\vec{x_n}, \vec{x}) + w_0 \quad \text{(approche duale)}$$

avec $t_n \in \{-1, 1\}$ (En fonction de $k(\vec{x_n}, \vec{x})$ matrice de gram et a_n optimisés avec l'entraînement)

Pour chercher les paramètres et hyperplans $(\vec{w}, w_0)$ qui maximise la marge on utilise SVM tq

$$\arg\max_{\vec{w}, w_0} \{\text{marge}(\vec{w}, w_0)\} = \arg\max_{\vec{w}, w_0} \left\{ \min_n \left(t_n \frac{y_{\vec{w}}(\vec{\phi}(x_n))}{\|\vec{w}\|} \right) \right\}$$

$$= \arg\max_{\vec{w}, w_0} \left\{ \frac{1}{\|\vec{w}\|} \min_n \left(t_n (\vec{w}^T \phi(\vec{x_n}) + w_0) \right) \right\}$$

↳ ∃ une infinité de solutions à ce problème.

donc il faut qu'on considère une contrainte sur les vecteurs support tq: $t_n(\vec{w}^T \vec{\phi}(\vec{x_n}) + w_0) \geq 1$

$$t_n y_{\vec{w}}(\vec{\phi}(\vec{x_n})) = t_n(\vec{w}^T \vec{\phi}(\vec{x_n}) + w_0) = 1$$

par l'approche duale $(*)$ seuls les vecteurs support vont figurer dans l'équation sur la contrainte:

$$\begin{cases} t_n y(\vec{x_n}) = 1 \text{ et } a_n > 0 \\ t_n y(\vec{x_n}) \geq 1 \text{ et } a_n = 0 \end{cases}$$

Or lorsque les données sont non séparables on doit introduire les variables de ressorts ξ_n qui correspondent aux violations des contraintes de marge.

→ si ξ_n est plus grande que 1 : $\xi_n > 1$ la donnée est mal classée.

$$\arg\min_{\vec{w}, w_0, \vec{\xi}} \left\{ \frac{1}{2} \|\vec{w}\|^2 \right\} + C \sum_{n=1}^{N} \xi_n \quad \text{tq } t_n y_n(\vec{\phi}(\vec{x_n})) \geq 1 - \xi_n \quad \forall n, \xi_n \geq 0$$

→ En tenant compte de la représentation duale.

$$\vec{\alpha}(\vec{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n t_m k(\vec{x_n}, \vec{x_m}) \quad \text{tq} \begin{cases} C \geq a_n \geq 0 \\ \sum_{n=1}^{N} a_n t_n = 0 \end{cases}$$

→ En tenant compte la représentation primale.

$$\arg\min_{\vec{w}, w_0, \vec{\xi}} \left\{ \frac{1}{2} \|\vec{w}\|^2 \right\} + C \sum_{n=1}^{N} \xi_n \quad \text{tq } t_n y_{\vec{w}}(\vec{\phi}(\vec{x_n})) \geq 1 - \xi_n \quad \forall n, \xi_n \geq 0$$

$$\Rightarrow \quad \xi_n \geq 1 - t_n y_n \vec{w}(\vec{\phi}(\vec{x_n}))$$

$$\arg\min_{\vec{w}, w_0} \underbrace{\frac{1}{2}\|\vec{w}\|^2}_{\text{terme de régularisation}} + C \sum_{n=1}^{N} \underbrace{\max\left(0, 1 - t_n y_{\vec{w}}(\vec{\phi}(\vec{x_n}))\right)}_{\text{notion de marge}}$$

fonction de perte (Hinge loss)

On retrouve donc une fct de perte du type Hinge loss avec un terme de régularisation.