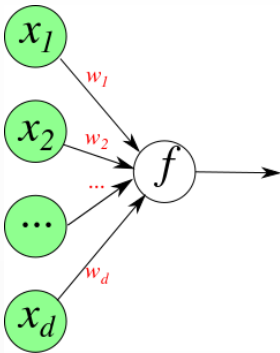
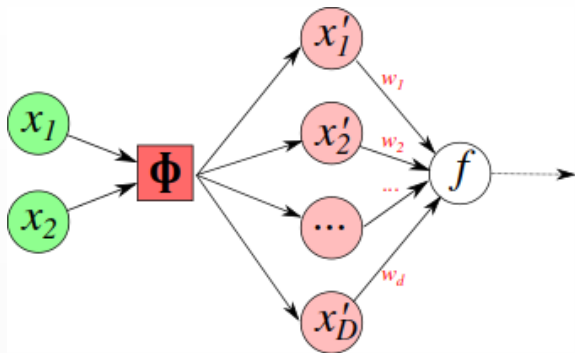


Introduction au SVM et aux méthodes à noyaux

Pascal Germain

2018





- 1 Classification binaire et marge du prédicteur
- 2 SVM à marge rigide **(1992)**
- 3 SVM à marge floue **(1995)**
- 4 Astuce du noyau
- 5 La régression de ridge à noyau
- 6 Approximation du noyau par *Random Fourier Features* **(2007)**
- 7 Références scikit-learn

Ensemble d'apprentissage :

$$S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\},$$

avec $\mathbf{x}_i \in \mathbb{R}^d$ et $y_i \in \{-1, +1\}$.

Prédicteur linéaire :

$$f_{\mathbf{w},b}(\mathbf{x}) = \text{sgn}[\mathbf{w} \cdot \mathbf{x} - b] = \begin{cases} +1 & \text{si } \mathbf{w} \cdot \mathbf{x} - b > 0 \\ -1 & \text{sinon.} \end{cases}$$

Note : $f_{\mathbf{w},b}(\mathbf{x}) = f_{\mathbf{w}',b'}(\mathbf{x})$ avec $\mathbf{w}' = c\mathbf{w}$ et $b' = cb$ pour tout $c > 0$.

Marge d'un prédicteur linéaire sur un exemple

Marge fonctionnelle du prédicteur $f_{\mathbf{w},b}$ sur l'exemple (\mathbf{x}, y) :

$$y (\mathbf{w} \cdot \mathbf{x} - b)$$

Marge géométrique du prédicteur $f_{\mathbf{w},b}$ sur l'exemple (\mathbf{x}, y) :

$$\frac{y (\mathbf{w} \cdot \mathbf{x} - b)}{\|\mathbf{w}\|}$$

Note 1 : Un exemple bien classifié ($f_{\mathbf{w},b}(\mathbf{x}) = y$) ssi sa marge est positive.

Note 2 : Avec $c > 0$, $\mathbf{w}' = c \mathbf{w}$ et $b' = c b$ les prédicteurs $f_{\mathbf{w},b}$ et $f_{\mathbf{w}',b'}$ possèdent la même marge géométrique.

Marge d'un prédicteur linéaire sur un ensemble d'apprentissage

Marge fonctionnelle du prédicteur $f_{\mathbf{w},b}$ sur l'ensemble S :

$$\min_{(\mathbf{x}_i, y_i) \in S} \left[y_i (\mathbf{w} \cdot \mathbf{x}_i - b) \right]$$

Marge géométrique du prédicteur $f_{\mathbf{w},b}$ sur l'ensemble S :

$$\min_{(\mathbf{x}_i, y_i) \in S} \left[\frac{y_i (\mathbf{w} \cdot \mathbf{x}_i - b)}{\|\mathbf{w}\|} \right]$$

Note : On dit qu'un ensemble S est linéairement séparable lorsqu'il existe un couple $(\mathbf{w}, b) \in \mathbb{R}^d \times \mathbb{R}$ tel que la marge sur l'ensemble S est positive.

SVM à marge rigide

Supposons que l'ensemble d'apprentissage S soit linéairement séparable.

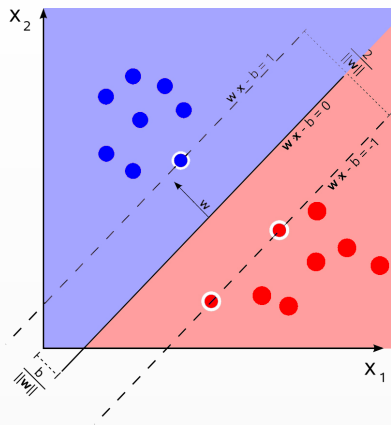
Le SVM à marge rigide trouve un prédicteur $f_{\mathbf{w},b}$ de marge géométrique maximale :

$$\max_{(\mathbf{w},b) \in \mathbb{R}^d \times \mathbb{R}} \left(\min_{(\mathbf{x}_i, y_i) \in S} \left[\frac{y_i (\mathbf{w} \cdot \mathbf{x}_i - b)}{\|\mathbf{w}\|} \right] \right)$$

Il existe une multitude de solutions (\mathbf{w}, b) à ce problème... Prenons la solution telle que la marge fonctionnelle sur l'ensemble S vaut 1, c'est-à-dire :

$$\min_{(\mathbf{x}_i, y_i) \in S} \left[\frac{y_i (\mathbf{w} \cdot \mathbf{x}_i - b)}{\|\mathbf{w}\|} \right] = \frac{1}{\|\mathbf{w}\|}$$

SVM à marge rigide



Étant donné :

$$S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}.$$

Minimiser : $\frac{1}{2} \|\mathbf{w}\|^2$

sous contraintes : $y_i (\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1$
pour $i = 1, \dots, n$.

On nomme *vecteurs de supports* les exemples dont la marge fonctionnelle est 1.

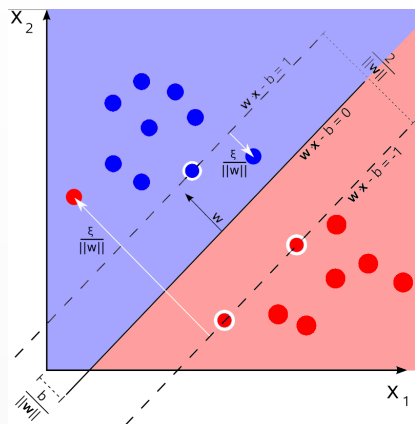
Pour d'adapter à la situation où S n'est pas linéairement séparable, on introduit des variables d'écart.

$$\xi_i = \max \{0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i - b)\}$$

Minimiser : $\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$

sous contraintes : $y_i (\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i$ et $\xi_i \geq 0$
pour $i = 1, \dots, n$.

SVM à marge floue



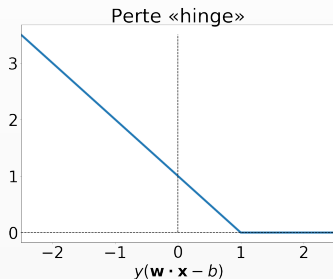
La valeur de chaque ξ_i s'interprète ainsi :

- $\xi_i > 1$: L'exemple est mal classifié.
- $0 < \xi_i < 1$: L'exemple est bien classifié, mais il est situé à l'intérieur de la marge du SVM.
- $\xi_i = 0$: L'exemple est bien classifié et il est situé à l'extérieur de la marge du SVM.

Minimiser :
$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \ell_{\text{hinge}}(f_{\mathbf{w},b}(\mathbf{x}_i), y_i)$$

avec

$$\ell_{\text{hinge}}(\hat{y}, y) = \max \{0, 1 - \hat{y} \times y\}$$



Le noyau linéaire

Étant donné :

$$S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}.$$

Représentons le vecteur $\mathbf{w} \in \mathbb{R}^d$ par un vecteur de variables duales $\boldsymbol{\alpha} \in \mathbb{R}^n$, tel que

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i.$$

Nous pouvons réécrire le prédicteur $f_{\mathbf{w},b}$:

$$\begin{aligned} f_{\mathbf{w},b}(\mathbf{x}) &= \operatorname{sgn}[\mathbf{w} \cdot \mathbf{x} + b] = \operatorname{sgn} \left[\sum_{i=1}^n \alpha_i \mathbf{x}_i \cdot \mathbf{x} + b \right] \\ &= \operatorname{sgn} \left[\sum_{i=1}^n \alpha_i k(\mathbf{x}_i, \mathbf{x}) + b \right], \end{aligned}$$

où $k(\mathbf{x}_i, \mathbf{x}) = \mathbf{x}_i \cdot \mathbf{x}$ est la fonction de **noyau linéaire**.

$$\begin{aligned}\|\mathbf{w}\|^2 &= \mathbf{w} \cdot \mathbf{w} = \left(\sum_{i=1}^n \alpha_i \mathbf{x}_i \right) \cdot \left(\sum_{i=1}^n \alpha_i \mathbf{x}_i \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbf{x}_i \cdot \mathbf{x}_j \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)\end{aligned}$$

Une fonction $k : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ est un *noyau* ssi il existe une transformation $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$ telle que

$$k(\mathbf{x}_i, \mathbf{x}) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x})$$

Autrement dit, un noyau calcule à un **produit scalaire dans un espace augmenté**.

Le SVM dans l'espace des noyaux

Définissons la matrice de noyau $\mathbf{K} \in \mathbb{R}^{n \times n}$ telle que

$$K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$$

Le problème d'optimisation en fonction de $\mathbf{w} \in \mathbb{R}^D$...

Minimiser : $\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$

sous contraintes : $y_i (\mathbf{w} \cdot \phi(\mathbf{x}_i) - b) \geq 1 - \xi_i$ et $\xi_i \geq 0 \quad \forall i \in \{1 \dots n\}$

...se réécrit en un problème d'optimisation en fonction de $\alpha \in \mathbb{R}^n$:

Minimiser : $\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K_{i,j} + C \sum_{i=1}^n \xi_i$

sous contraintes : $y_i \left(\sum_{j=1}^n \alpha_j K_{i,j} - b \right) \geq 1 - \xi_i$ et $\xi_i \geq 0 \quad \forall i \in \{1 \dots n\}$

Théorème du représentant

Soit un noyau $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$, une fonction de perte quelconque $\ell : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, un ensemble d'apprentissage S de n observations et un paramètre de régularisation $\lambda > 0$.

La solution au problème

$$\mathbf{w}^* = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^D} \left[\sum_{i=1}^n \ell(\mathbf{w} \cdot \phi(\mathbf{x}_i), y_i) + \lambda \|\mathbf{w}\|^2 \right]$$

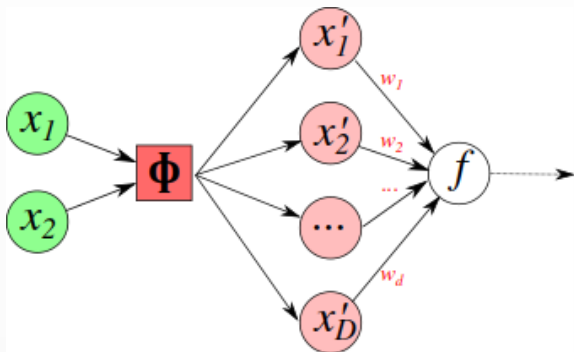
peut s'écrire

$$\mathbf{w}^* = \sum_{j=1}^n \alpha_j \phi(\mathbf{x}_j)$$

où α est une solution

$$\min_{\alpha \in \mathbb{R}^d} \left[\sum_{i=1}^n \ell \left(\sum_{j=1}^n \alpha_j k(\mathbf{x}_i, \mathbf{x}_j), y_i \right) + \lambda \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j), y_i \right]$$

Par l'*astuce du noyau*, la représentation $\mathbf{x}' \in \mathbb{R}^D$
est utilisée de manière implicite !

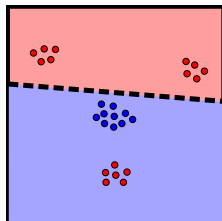


Démo : <https://www.youtube.com/watch?v=3liCbRZPrZA>

Quelques exemples de noyaux

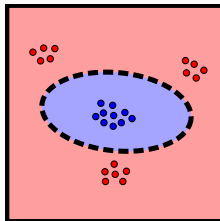
Noyau linéaire

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$$



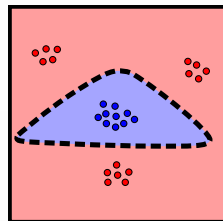
Noyau polynomial
(degré 2)

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^2$$



Noyau gaussien

$$k(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|}$$



Le noyau polynomial

$$k_{\text{poly}(p)}(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x} \cdot \mathbf{x}')^p \quad \text{avec } p \in \mathbb{N}^+$$

Montrons que le noyau polynomial de degré 2 correspond bien à un produit scalaire. On considère que $x_0 = x'_0 = 1$.

$$\begin{aligned} k_{\text{poly}(2)}(\mathbf{x}, \mathbf{x}') &= (1 + \mathbf{x} \cdot \mathbf{x}')^2 \\ &= (x_0 x'_0 + x_1 x'_1 + x_2 x'_2 + \dots + x_d x'_d)^2 \\ &= \sum_{i=0}^d \sum_{j=0}^d (x_i x'_i)(x_j x'_j) \\ &= \sum_{i=0}^d \sum_{j=0}^d (x_i x_j)(x'_i x'_j) \\ &= \phi_{\text{poly}(2)}(\mathbf{x}) \cdot \phi_{\text{poly}(2)}(\mathbf{x}') \end{aligned}$$

où

$$\phi_{\text{poly}(2)}(\mathbf{x}) = (x_0 x_0, x_0 x_1, \dots, x_0 x_d, \dots, x_d x_0, x_d x_1, \dots, x_d x_d) \in \mathbb{R}^{(d+1)^2}.$$

$$k_{\text{poly}(p)}(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x} \cdot \mathbf{x}')^p \quad \text{avec } p \in \mathbb{N}^*$$

Le résultat précédent se généralise à

$$k_{\text{poly}(p)}(\mathbf{x}, \mathbf{x}') = \phi_{\text{poly}(p)}(\mathbf{x}) \cdot \phi_{\text{poly}(p)}(\mathbf{x}')$$

où chaque élément du vecteur $\phi_{\text{poly}(p)}(\mathbf{x}) \in \mathbb{R}^{(d+1)^p}$ est de la forme :

$$x_{i_1} x_{i_2} \cdots x_{i_p} \quad \text{pour } \{i_1, \dots, i_p\} \in \{0, 1, 2, \dots, d\}^p$$

Le noyau gaussien (ou *RBF*, pour *Radial Basis Function*)

$$k_\gamma(\mathbf{x}, \mathbf{x}') = e^{-\gamma\|\mathbf{x}-\mathbf{x}'\|^2} \quad \text{ou } k_{\sigma^2}(\mathbf{x}, \mathbf{x}') = e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}} \quad \text{avec } \gamma = \frac{1}{2\sigma^2} > 0$$

$$\begin{aligned} k_\gamma(\mathbf{x}, \mathbf{x}') &= e^{-\gamma\|\mathbf{x}-\mathbf{x}'\|^2} \\ &= e^{-\gamma(\|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2 - 2\mathbf{x}\cdot\mathbf{x}')} \\ &= e^{-\gamma\|\mathbf{x}\|^2} e^{-\gamma\|\mathbf{x}'\|^2} e^{\gamma 2\mathbf{x}\cdot\mathbf{x}'} \\ &= e^{-\gamma\|\mathbf{x}\|^2} e^{-\gamma\|\mathbf{x}'\|^2} \sum_{n=0}^{\infty} \frac{(\gamma 2\mathbf{x}\cdot\mathbf{x}')^n}{n!} \\ &= e^{-\gamma\|\mathbf{x}\|^2} e^{-\gamma\|\mathbf{x}'\|^2} \sum_{n=0}^{\infty} \frac{(2\gamma)^n}{n!} k_{\text{poly}(n)}(\mathbf{x}, \mathbf{x}') \\ &= \phi_\gamma(\mathbf{x}) \cdot \phi_\gamma(\mathbf{x}') \end{aligned}$$

où $\phi_\gamma(\mathbf{x}) \in \mathbb{R}^\infty$ est la concaténation (infinie) des vecteurs

$$\sqrt{\frac{(2\gamma)^n}{n!}} e^{-\gamma\|\mathbf{x}\|^2} \phi_{\text{poly}(n)}(\mathbf{x}) \quad \text{pour tout } n \in \mathbb{N}$$

On peut aussi montrer qu'une fonction k est un noyau en montrant que la matrice de noyau \mathbf{K} est toujours positive semi-définie.

Autrement dit, pour tout ensemble d'apprentissage S de n observations et pour tout $\mathbf{z} \in \mathbb{R}^n$, on a

$$\sum_{i=1}^n \sum_{j=1}^n z_i z_j k(\mathbf{x}_i, \mathbf{x}_j) > 0$$

La régression de ridge à noyau

Considérons un problème de régression, où $(\mathbf{x}, y) \in \mathbb{R}^n \times \mathbb{R}$,
et un noyau $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$.

$$f_{\mathbf{w}}(\phi(\mathbf{x})) = \mathbf{w} \cdot \phi(\mathbf{x}) = \sum_{i=1}^n \alpha_i k(\mathbf{x}, \mathbf{x}_i)$$

Le problème de la régression logistique en fonction de $\mathbf{w} \in \mathbb{R}^D$...

$$\min_{\mathbf{w} \in \mathbb{R}^D} \left[\frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{w} \cdot \phi(\mathbf{x}_i))^2 + \lambda \|\mathbf{w}\|^2 \right]$$

...s'exprime en fonction des variables $\boldsymbol{\alpha} \in \mathbb{R}^n$

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \left[\frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^n \alpha_j K_{i,j} \right)^2 + \lambda \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K_{i,j} \right]$$

La régression de ridge à noyau

Le problème de la régression logistique en fonction de $\mathbf{w} \in \mathbb{R}^D$...

$$\operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^D} \left[\frac{1}{n} \|\mathbf{y} - \Phi \mathbf{w}\|^2 + \lambda \|\mathbf{w}\|^2 \right] = (\Phi^T \Phi + \lambda n \mathbf{I})^{-1} \Phi^T \mathbf{y}$$

...s'exprime en fonction des variables $\boldsymbol{\alpha} \in \mathbb{R}^n$

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \left[\frac{1}{n} \|\mathbf{y} - \mathbf{K} \boldsymbol{\alpha}\|^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} \right] = (\mathbf{K} + \lambda n \mathbf{I})^{-1} \mathbf{y}$$

Random Fourier Features (RFF)

Considérons un noyau *invariant par translation* $k(\boldsymbol{\delta}) = k(\mathbf{x} - \mathbf{x}') = k(\mathbf{x}, \mathbf{x}')$.

Soit $p(\boldsymbol{\omega})$ la transformée de Fourier de $k(\boldsymbol{\delta})$,

$$p(\boldsymbol{\omega}) = \frac{1}{2\pi} \int_{\mathbb{R}^d} k(\boldsymbol{\delta}) e^{-i\boldsymbol{\omega} \cdot \boldsymbol{\delta}} d\boldsymbol{\delta},$$

Exemple du noyau gaussien

Le noyau gaussien (RBF) correspond à

$$k_{\sigma}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}'\|^2\right).$$

La transformée de Fourier associée est

$$p_{\sigma}(\boldsymbol{\omega}) = \left(\frac{\sigma^2}{2\pi}\right)^{\frac{d}{2}} e^{-\frac{1}{2}\sigma^2 \|\boldsymbol{\omega}\|^2} = \mathcal{N}(\boldsymbol{\omega}; \mathbf{0}, \frac{1}{\sigma^2} \mathbf{I}).$$

Random Fourier Features (RFF)

Considérons un noyau *invariant par translation* $k(\boldsymbol{\delta}) = k(\mathbf{x} - \mathbf{x}') = k(\mathbf{x}, \mathbf{x}')$.

Soit $p(\boldsymbol{\omega})$ la transformée de Fourier de $k(\boldsymbol{\delta})$,

$$p(\boldsymbol{\omega}) = \frac{1}{2\pi} \int_{\mathbb{R}^d} k(\boldsymbol{\delta}) e^{-i\boldsymbol{\omega} \cdot \boldsymbol{\delta}} d\boldsymbol{\delta},$$

Alors,

$$\begin{aligned} k(\mathbf{x} - \mathbf{x}') &= \int_{\mathbb{R}^d} p(\boldsymbol{\omega}) e^{i\boldsymbol{\omega} \cdot (\mathbf{x} - \mathbf{x}')} d\boldsymbol{\omega} \\ &= \mathbf{E}_{\boldsymbol{\omega} \sim p} e^{i\boldsymbol{\omega} \cdot (\mathbf{x} - \mathbf{x}')} \\ &= \mathbf{E}_{\boldsymbol{\omega} \sim p} \left[\cos(\boldsymbol{\omega} \cdot (\mathbf{x} - \mathbf{x}')) + i \sin(\boldsymbol{\omega} \cdot (\mathbf{x} - \mathbf{x}')) \right] \\ &= \mathbf{E}_{\boldsymbol{\omega} \sim p} \cos(\boldsymbol{\omega} \cdot (\mathbf{x} - \mathbf{x}')) + i \mathbf{E}_{\boldsymbol{\omega} \sim p} \sin(\boldsymbol{\omega} \cdot (\mathbf{x} - \mathbf{x}')) \\ &= \mathbf{E}_{\boldsymbol{\omega} \sim p} \cos(\boldsymbol{\omega} \cdot (\mathbf{x} - \mathbf{x}')). \end{aligned}$$

Soit $\omega \sim p$ et $b \sim \text{UNIFORME}([0, 2\pi])$:

$$z_{\omega,b}(\mathbf{x}) = \sqrt{2} \cos(\omega \cdot \mathbf{x} + b),$$

La variable aléatoire $z_{\omega,b}(\mathbf{x}) \cdot z_{\omega,b}(\mathbf{x}')$ est un estimateur non biaisé de $k(\mathbf{x} - \mathbf{x}')$.

$$\begin{aligned} & \mathbf{E}_{\omega \sim p} \mathbf{E}_{b \sim u} z_{\omega,b}(\mathbf{x}) \cdot z_{\omega,b}(\mathbf{x}') \\ &= \mathbf{E}_{\omega \sim p} \mathbf{E}_{b \sim u} \left[2 \cos(\omega \cdot \mathbf{x} + b) \cos(\omega \cdot \mathbf{x}' + b) \right] \\ &= \mathbf{E}_{\omega \sim p} \mathbf{E}_{b \sim u} \left[\cos(\omega \cdot \mathbf{x} + b - \omega \cdot \mathbf{x}' + b) + \cos(\omega \cdot \mathbf{x} + b + \omega \cdot \mathbf{x}' + b) \right] \\ &= \mathbf{E}_{\omega \sim p} \mathbf{E}_{b \sim u} \left[\cos(\omega \cdot (\mathbf{x} - \mathbf{x}')) + \cos(\omega \cdot (\mathbf{x} + \mathbf{x}') + 2b) \right] \\ &= \mathbf{E}_{\omega \sim p} \left[\cos(\omega \cdot (\mathbf{x} - \mathbf{x}')) + \int_0^{2\pi} \frac{1}{2\pi} \cos(\omega \cdot (\mathbf{x} + \mathbf{x}') + 2b) db \right] \\ &= \mathbf{E}_{\omega \sim p} \cos(\omega \cdot (\mathbf{x} - \mathbf{x}')) + 0 \\ &= k(\mathbf{x} - \mathbf{x}'). \end{aligned}$$

Procédure d'apprentissage

- Échantillonner D points selon p : $\omega_1, \omega_2, \dots, \omega_D$.
- Échantillonner D points selon u : b_1, b_2, \dots, b_D .
- Transformer chaque exemple d'apprentissage $\mathbf{x}_j \in \mathbb{R}^d$ en un nouveau vecteur $\phi(\mathbf{x}_j) \in \mathbb{R}^D$:

$$\phi(\mathbf{x}_j) = \frac{1}{\sqrt{D}} \left(\cos(\omega_1 \cdot \mathbf{x}_j + b_1), \cos(\omega_2 \cdot \mathbf{x}_j + b_2), \dots, \cos(\omega_D \cdot \mathbf{x}_j + b_D) \right).$$

- Apprendre un prédicteur linéaire $\mathbf{w} \in \mathbb{R}^D$ sur l'ensemble transformé $\{(\phi(\mathbf{x}_j), y_j)\}_{j=1}^n$:

$$f(\mathbf{x}) = \sum_{j=1}^D w_j \phi_j(\mathbf{x}),$$

La magie opère !

Quand D est "suffisamment grand", on obtient $k(\mathbf{x} - \mathbf{x}') \approx \phi(\mathbf{x}) \cdot \phi(\mathbf{x}')$.

SVM Lineaire

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html>

SVM à noyau

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

Régression de ridge à noyau

https://scikit-learn.org/stable/modules/generated/sklearn.kernel_ridge.KernelRidge.html

RFF (noyau gaussien)

https://scikit-learn.org/stable/modules/generated/sklearn.kernel_approximation.RBFSampler.html