Rappels (supervisé)

- Regression linéaire
  - ↳ Descente de gradient
  - Equation Normales
  - Regularisation
  - Equilibre biais-variance
  - Motivations statistiques (MLE / MAP)

- Classification
  - → Modèle linéaire
  - → Extension à plusieurs classes (1 vs 1, 1 vs tous
    Discriminant à plusieurs classes)
  - → Regression logistique
  - Perceptron

→ Reseaux de neurones (Formulation générale + Back propagation)

Aujourd'hui : Astuce du noyau / Méthode à noyau

$$l(\beta) = \frac{1}{N} \sum_{i=1}^{N} \left(t^{(i)} - \langle \beta, \tilde{x}^{(i)} \rangle \right)^2 \qquad \tilde{x}^{(i)} = [1, x_1^{(i)}, \dots, x_D^{(i)}]$$

$\tilde{\Phi}(x^{(i)})$ = vecteur de caractéristiques supplementaires
(ex : polynomiales)

$$\beta \leftarrow \beta + \frac{2\eta}{N} \sum_{i=1}^{N} \left(t^{(i)} - \langle \beta, \tilde{\Phi}(x^{(i)}) \rangle \right) \tilde{\Phi}(x^{(i)})$$

ex

$\tilde{x}^{(i)} \in \mathbb{R}^D$ $\longrightarrow$ si $\tilde{\Phi}(x^{(i)})$ est composé de caractéristiques polynomiales de degré au max 3

$$\tilde{\Phi}(x^{(i)}) = [1, x_1, x_2, \dots, x_D, x_1^2, x_1 x_2, \dots, x_D^2, x_1^3, x_1^2 x_2, x_1^2 x_3, \dots, x_D^3]$$

$$\dim \tilde{\Phi}(x^{(i)}) = 1 + D + D^2 + D^3$$

supposons qu'on puisse écrire β comme une combinaison linéaire

$$\beta = \sum_{j=1}^{N} \lambda_j \tilde{\Phi}(x^{(j)}) \leftarrow$$

toujours vrai pour $\beta = 0$

$$\beta^{(k+1)} = \sum_{j=1}^{N} \lambda_j \tilde{\Phi}(x^{(j)}) + \frac{2\eta}{N} \sum_{j=1}^{N} \left(t^{(j)} - \sum_{i=1}^{N} \lambda_i \tilde{\Phi}(x^{(i)})^T \tilde{\Phi}(x^{(j)}) \right) \tilde{\Phi}(x^{(j)})$$

$$= \sum_{j=1}^{N} \tilde{\Phi}(x^{(j)}) \left[\lambda_j + \frac{2\eta}{N} \left(t^{(j)} - \sum_{i=1}^{N} \lambda_i \tilde{\Phi}(x^{(i)})^T \tilde{\Phi}(x^{(j)}) \right) \right]$$

$\underline{N \lll D^d}$

$$\boxed{\lambda_j^{(k+1)} = \lambda_j^{(k)} + \frac{2\eta}{N}\left(t^{(j)} - \sum_{i=1}^{N} \lambda_i^{(k)} \underbrace{\tilde{\Phi}(x^{(i)})^T \tilde{\Phi}(x^{(j)})}\right)}$$

$\langle \tilde{\phi}(x), \tilde{\phi}(z) \rangle$ pour $\tilde{\phi}(x) = [1, x_1, x_2, \dots, x_D, \overbrace{x_1^2, x_1 x_2, \dots, x_D^2}$
$x_1^3, x_1 x_2^2, \dots x_D^3]$

$$= 1 + \sum_{i=1}^{D} x_i z_i + \sum_{i,j} x_i x_j z_i z_j + \sum_{i,j,k} x_i x_j x_k z_i z_j z_k$$

$$= 1 + \langle x, z \rangle + \langle x, z \rangle^2 + \langle x, z \rangle^3$$

$$\lambda_j^{(k+1)} \leftarrow \lambda_j^{(k)} + \frac{2\eta}{N}\left(t^{(j)} - \sum_{i=1}^{N} \lambda_i^{(k)} K(i,j) \right) \leftarrow$$

$K(i,j)$ = matrice noyeau dont l'entrée (i,j) correspond au produit scalaire entre les vecteurs $\tilde{\phi}(x^{(i)})$ et $\tilde{\phi}(x^{(j)})$ (pas nécessairement connus explicitement)

Etant donné K, quelles sont les propriétés qui en font une "bonne" matrice pour l'apprentissage ?

Condition #1 : Symétrie

Condition # 2 : K semidefinie positive $K \geq 0$

$z^T K z \geq 0 \iff K = \Phi \Phi^T \leftarrow$
$\forall z$

suppose $K(i,j) = \tilde{\Phi}(x^{(i)})^T \tilde{\Phi}(x^{(j)})$

$$z^T K z = \sum_{i,j} z_i K(i,j) z_j = \sum_{i,j} z_i \overbrace{\tilde{\Phi}(x^{(i)})^T \tilde{\Phi}(x^{(j)})} z_j$$

$$= \sum_{i,j} z_i \sum_k \tilde{\Phi}_k(x^{(i)}) \tilde{\Phi}_k(x^{(j)}) z_j$$

$$= \sum_k \sum_i z_i \tilde{\Phi}_k(x^{(i)}) \sum_j z_j \tilde{\Phi}(x^{(j)})$$

$$= \sum_k \left(\sum_i z_i \tilde{\Phi}_k(x^{(i)}) \right)^2 \geq 0$$

Théorème de Mercer

K est un noyau valide ssi symétrique et semi-défini positif

Ex $K(x,z) = (x^T z)^2$

$$= \left(\sum_{i=1}^{d} x_i z_i\right)^2 = \sum_{i,j} x_i x_j z_i z_j$$

$$= \tilde{\phi}(x)^T \tilde{\phi}(z) \quad \text{où}$$

$$\tilde{\phi}(x) = [x_1^2, x_1 x_2, \dots, x_d^2]$$

$$K(x,z) = (x^T z + c)^2 \qquad c \geq 0$$

$$= c^2 + \left(\sum_i x_i z_i\right)^2 + 2c \sum_i x_i z_i$$

$$= c^2 + \sum_i \sum_j x_i x_j z_i z_j + 2c \sum_i x_i z_i = \tilde{\phi}(x)^T \tilde{\phi}(z)$$

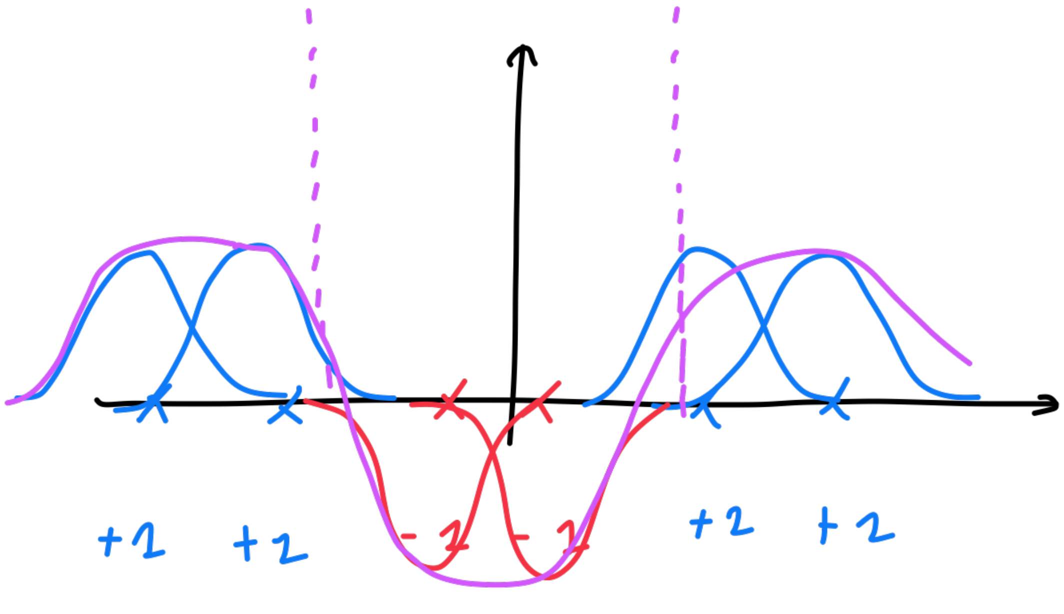
$$\tilde{\phi}(x) = [c, x_1^2, x_1 x_2, \dots, x_D^2, \sqrt{2c}\,(x_1, x_2, \dots, x_D)]$$

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{\sigma^2}\right)$$

$$y(x) = \beta^T \tilde{\phi}(x)$$

$$= \sum_{i=1}^{N} \lambda_i^\infty \, \tilde{\phi}(x^{(i)})^T \tilde{\phi}(x)$$

$$= \sum_{i=1}^{N} \lambda_i^\infty K(x^{(i)}, x) = \sum_{i=1}^{N} \lambda_i^\infty \exp\left(-\frac{\|x^{(i)} - x\|^2}{\sigma^2}\right)$$


+2
+2
-2
-2
+2
+2