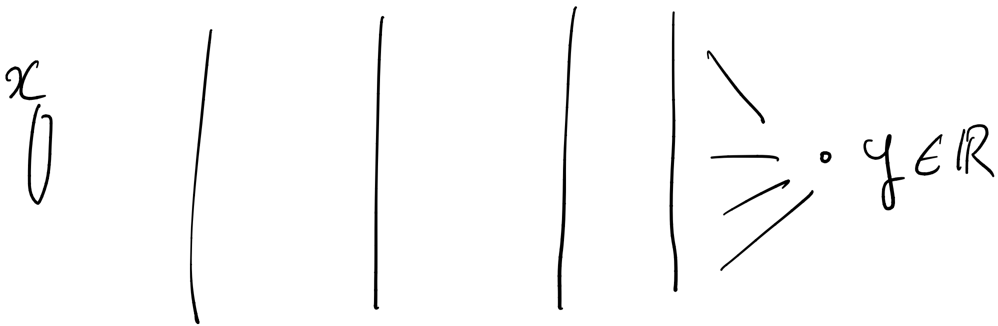


$$y = \beta^T x^L$$
$$= \underbrace{\beta^T \phi(\theta, x)}$$

$$x^L = \phi(\theta, x)$$

$$x \xrightarrow{W^1} z^1 = W^1 x \xrightarrow{\sigma} x^2 \rightarrow \cdots$$

à θ fixé

$\hookrightarrow$ RKHS

$$k(x, y) = \langle \phi(\theta, x), \phi(\theta, y) \rangle_H$$

$$\mathcal{X} \xrightarrow{\phi(\theta, \cdot)} H$$

$$x \rightsquigarrow \phi(\theta, x)$$

Objectif :

1) comprendre origine initialisation standard

2) voir que avec initialisation standard : NN = RKHS associé à $\nabla_\theta \phi(\theta_0, x)$ NTK

$\Rightarrow$ pas de feature learning

3) voir quelle param / init. pour avoir feature learning dans NN

4?) quel gain quand il y a feature learning. $L = 1$

Déroulé :

1) RKHS (lecture) → RKHS
→ representant
→ geometrie (SVM)

2) on va regarder 1 NN ultra-simple

→ init. standard

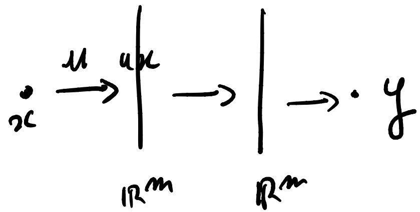
→ NTK pour lui

→ généralisation du calcul.

3) Thm : marche cv. (Chizat & Bach) (à voir)

4) comprendre la bonne param/init dans l'exemple du 2)

Reseau jouet : $\cdot\ L=2$ $\cdot\ x, y \in \mathbb{R}$ $\cdot\ \sigma(x) = x$



$$\phi(\theta, x) = \frac{1}{m^\alpha}\underset{\mathbb{R}^m}{\beta^T}\left(\frac{1}{m^\gamma}\underset{\mathbb{R}^{m\times m}}{A}\right)\underset{\mathbb{R}^m}{u}\,x$$

1) Initialisation : $u_i^0, A_{ij}^0, \beta_{ij}^0 \overset{iid}{\sim} \mathcal{N}(0,1)$

a) $\|u\|^2 \underset{m\to\infty}{\sim}$?

b) conditionnellement à u^0

quelle est la loi de $\frac{1}{m^\gamma} A^0 u^0 x$?

En déduire la loi de $\frac{1}{m^\gamma} A^0 u^0 x$

et $\left\|\frac{1}{m^{\gamma}} A^0 u^0 x\right\|^2 \underset{m\to\infty}{\sim}$?

comment choisir γ ?

c) loi de $\frac{1}{m^\alpha}\beta^T x^2$ cond^n à x^2

où $x^2 = \frac{1}{m^\gamma} A^0 u^0 x$.

en déduire la loi

d) comment choisir α et γ pour

avoir $\phi(\underbrace{(\beta^0, A^0, u^0)}_{=\theta^0}, x) \xrightarrow[m\to\infty]{} \begin{matrix}\neq 0 \\ \neq \pm\infty\end{matrix}$

2) Apprentissage :

a) $\nabla_\beta \phi$, $\nabla_A \phi$, $\nabla_u \phi$

2) Soit $F : \mathbb{R} \to \mathbb{R}$ avec $F'(0) \neq 0$ et

$$\Delta F = F(\phi(\theta^1, x)) - F(\phi(\theta^0, x))$$

1 pas de GD de paramètre $\eta = m^a$

avec l'approx. $\hookrightarrow \theta^1 = \theta^0 - \eta \nabla_\theta F(\phi(\theta_0, x))$

$$\Delta F \simeq \langle \Delta \theta, \nabla_\theta F(\phi(\theta_0, x)) \rangle$$

comment choisir η pour avoir

$\Delta F \nearrow \neq 0$
$\searrow \neq \pm \infty$

3) quel est l'ordre de grandeur de $\Delta \beta$, ΔA, Δu pour le choix de η ?

<u>ΔA ?</u> $\nearrow \Delta A_{ij}$ $\quad + \ |\Delta A|_{op}$
$\searrow \|\Delta A\|_F$

Rmq : poly MIA2 sect 7.3
{donne 1 info sur $|A^0|_{op}$