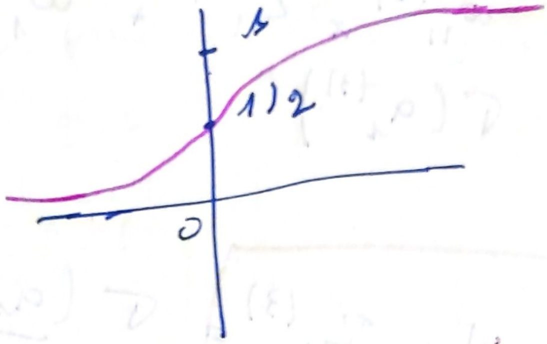
Exercices de Révision Sys

Q1:

1- $\sigma(x) = \dfrac{1}{1+e^{-x}}$



2-

~~y(x)~~ Feed forward:

- couche 1

$$\begin{cases} a_1^{(1)} = x_1 w_{11}^{(1)} + w_{10}^{(1)} \\ z_1^{(1)} = \sigma(a_1^{(1)}) \end{cases}$$

$$\begin{cases} a_2^{(1)} = x_2 w_{21}^{(1)} + w_{20}^{(1)} \\ z_2^{(1)} = \sigma(a_2^{(1)}) \end{cases}$$

$$a_3^{(1)} = x_3 w_{31}^{(1)} + w_{30}^{(1)}$$

$$z_3^{(1)} = \sigma(a_3^{(1)})$$

- couche 2:

$$\begin{cases} a_1^{(2)} = w_{11}^{(2)} \times z_1^{(1)} + w_{12}^{(2)} \times z_2^{(1)} + w_{10}^{(2)} \\ z_1^{(2)} = \sigma(a_1^{(2)}) \end{cases}$$

$$\begin{cases} a_2^{(2)} = w_{12}^{(2)} \times z_2^{(1)} + w_{22}^{(2)} \times z_3^{(1)} + w_{20}^{(2)} \\ z_2^{(2)} = \sigma(a_2^{(2)}) \end{cases}$$

couche 3 :
→ Sortie

$a_1^{(3)} = w_{11}^{(3)} \times z_1^{(2)} + w_{12} . z_2^{(2)} + w_{10}^{(3)}$

$z_1^{(3)} = \sigma(a_1^{(3)})$

$\boxed{y(w.x) = z_1^{(3)}} = \boxed{\sigma(\underbrace{a_1^{(3)}}_{a\,out})}$ ✓

3 - Back prabagation

$\ell(B) = -\sum_{i=1}^{N} \log t \; y(w.x) + (1-t) \log(1 - y(w.x))$

$y(w.x) = \sigma(a_1^{(3)})$

$\boxed{\frac{\partial \ell(B)}{\partial a_1^{(3)}} = \sigma(a_1^{(3)}) - t}$ ✓

- dérivation en chaine

$\frac{\partial a_1^{(3)}}{\partial w_{11}^{(1)}} = \frac{\partial a_1^{(3)}}{\partial z_1^{2}} \times \frac{\partial z_1^{(2)}}{\partial a_1^{(2)}} \times \frac{\partial a_1^{(2)}}{\partial z_1^{(1)}} \times \frac{\partial z_1^{(1)}}{\partial a_1^{(1)}} \times \frac{\partial a_1^{(1)}}{\partial w_{11}^{(1)}}$

$= w_{11}^{(3)} \times \sigma'(a_1^{(2)}) \times w_{11}^{(2)} \; \sigma'(a_1^{(1)}) \times x_1$

$\frac{\partial \ell}{\partial w_{11}^{(1)}} = \frac{\partial \ell}{\partial a_1^{(3)}} \times \frac{\partial a_1^{(3)}}{\partial w_{11}^{(1)}} =$

$= (\sigma(a_1^{(3)}) - t) \times w_{11}^{(3)} \times \sigma'(a_1^{(2)}) \times w_{11}^{(2)} \times \sigma'(a_1^{(1)}) \times x_1$

7

Question 2 :

1. dans le graphe, on peut diviser les données en 2 parties sachant $x_2 = -2$

on sait que $y(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

et on a $x_2 + 2 = 0$

donc
$$2 + 0 \times x_1 + x_2 = 0$$
$$\beta_0 + \beta_1 x_2 + \beta_2 x_2 = 0$$

par identification on trouve

$$\begin{cases} \beta_0 = 2 \\ \beta_1 = 0 \\ \beta_2 = 1 \end{cases} \Rightarrow \beta(2, 0, 1)$$

2 > Minimiser la fct coût : Régularisation Ridge

on a : MSE
$$l(\beta) = \frac{1}{N} \sum_{i=1}^{N} (t^i - h(\beta))^2 + \lambda \sum_{j=1}^{2} \beta_j^2$$

si nous utilisons une fonction de coût type Ridge les valeur β_0 et β_1 ne changerons pas car la pénalité Ridge n'affecte pas β_0 et aussi β_1 est le plus petit possible. ~~Pour β_2 ...~~
en ce qui conserne β_2 il va décroitre en fonction du coefficient de pénalité λ : $\beta_2 = 1 - \delta$

$\delta \geq 0$

Question 4 :

1. $\varepsilon \sim N(0, \sigma^2)$

$$\varepsilon = t - \hat{t}$$

$$\hat{t} = \beta_0 + \beta_1 x_1 + \dots \beta_d x_d$$

$$t = \hat{t} + \sigma$$

$$= \beta_0 + \beta^T x + \varepsilon$$

on a $\varepsilon \sim (0, \sigma^2)$ $-(\varepsilon - u)^2, 0$

$$p(t \mid x, \beta) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-\varepsilon^2}{2\sigma^2}\right)$$

$$\varepsilon = t - \hat{t}$$

$$= t - \beta_0 + \beta^T x$$

$$p(t \mid x, \beta) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp -\left(\frac{t - \beta_0 \beta^T x)^2}{2\sigma^2}\right)$$

$$\mathcal{L}(\beta) = \prod_{i=1}^{N} p(t = t^{(i)} \mid x^i, \beta)$$

$$= \prod \left(\frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-1}{2\sigma^2}\left(t - \beta_0 - \beta^T x^i\right)^2\right)\right)$$

$$\boxed{\log \mathcal{L}(\beta) = \sum_{i=1}^{N}\left(\log\left(\frac{1}{\sqrt{2\pi\sigma}}\right) + \underbrace{\left(\frac{-1}{2\sigma^2}\right)\left(t - \beta_0 - \beta^T x^i\right)^2}_{A}\right)}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}}$$

≥ 0

9

2. $\dfrac{\partial \log \mathcal{L}(\beta)}{\partial \sigma} = 0$

$$= \sum \left(\frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)'}{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^2} + \frac{-1 \times (-2\sigma^2)'}{(2\sigma^2)^2} \times A \right)$$

$$\left(\sqrt{f(x)}\right)' = \frac{f'(x)}{2\sqrt{f(x)}}$$

$$\log(f(x))' = \frac{f'(x)}{f(x)}$$

$$= \sum \left(\frac{-\frac{\left(\sqrt{2\pi\sigma^2}\right)'}{\left(\sqrt{2\pi\sigma^2}\right)^2}}{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^2} + \frac{4\sigma}{(2\sigma^2)^2} \times A \right)$$

$$= \sum \left(-\frac{(2\pi\sigma^2)'}{2\sqrt{2\pi\sigma^2}} + \frac{1}{\sigma^3} \times A \right)$$

$$\frac{\partial \log \mathcal{L}(\beta)}{\partial \sigma} = \sum \left(-\frac{4\pi\sigma}{2\sqrt{2\pi\sigma^2}} + \frac{1}{\sigma^3} \times (t - \beta_0 - \beta^T x^i)^2 \right)$$

$$= \sum \left(\frac{-2\pi}{\sqrt{2\pi}} + \frac{1}{\sigma^3} \times (t - \beta_0 - \beta^T x^i)^2 \right) \quad (1)$$

$$= \sum \left(-\sqrt{2\pi} + \frac{1}{\sigma^3} \times (t - \beta_0 - \beta^T x^i)^2 \right)$$

$$\frac{\partial^2 \log \mathcal{L}(\beta)}{\partial \sigma^2} = \sum \left(0 + \frac{-(\sigma^3)'}{(\sigma^3)^2} \times (t - \beta_0 - \beta^T x^i)^2 \right)$$

$$= \sum \left(\frac{-3\sigma^2}{\sigma^6} \times (t - \beta_0 - \beta^T x^i)^2 \right)$$

$$= \sum \left(\frac{-3}{\sigma^4} \times (t - \beta_0 - \beta^T x^i)^2 \right) = 0 \quad (2)$$

résoudre $\frac{\partial \log(\mathcal{L}(\beta))}{\partial \sigma} = 0$

$$\Rightarrow -\sqrt{2\pi} + \frac{1}{\sigma^3}(t - \beta_0 - \beta^T x)^2 = 0$$

$$\Rightarrow \frac{1}{\sigma^3} = \frac{-\sqrt{2\pi}}{(t - \beta_0 - \beta^T x)^2}$$

$$\Rightarrow \sigma^3 = -\frac{(t - \beta_0 - \beta^T x)^2}{\sqrt{2\pi}}$$

$$\boxed{\Rightarrow \sigma = \left(-\frac{1}{\sqrt{2\pi}}(t - \beta_0 - \beta^T x)^2\right)^{1/3}}$$

Question 5.2)

$\operatorname{argmin} \frac{1}{2}\sum (t^i - \beta x^{(i)})^2$ avec

$$\begin{cases} \frac{\partial \mathcal{L}(\beta)}{\partial \beta_0} = \sum (t^i - \beta x^{(i)}) = 0 \\ \frac{\partial \mathcal{L}(\beta)}{\partial \beta_j} = \sum (t^i - \beta x^{(i)}) x^{(i)} = 0 \\ j \neq 0 \end{cases}$$

5.3

Question 13 :

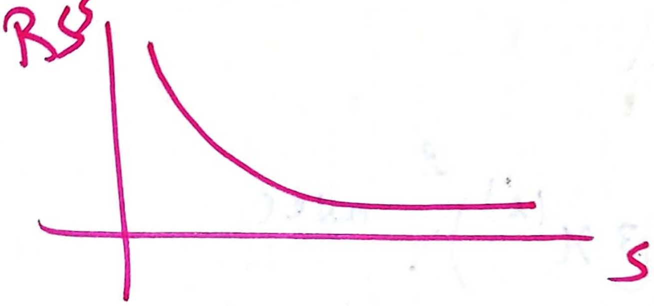
1 - $\sum_{i=1} (t^{(i)} - \beta_0 - \sum \beta_j x_{ij})^2 \qquad \sum |\beta_j| \leq s$

Si $s \uparrow \rightsquigarrow$ RSS

$\lambda |\beta| \leq s$

$\begin{cases} \text{Si } \lambda \uparrow \sim \beta \downarrow \Rightarrow \text{forte pénalisation} \sim RSS \uparrow \\ \text{Si } \lambda \downarrow \sim \beta \uparrow \Rightarrow \text{faible pénalisation} \rightsquigarrow RSS \downarrow \end{cases}$

- Faux, RSS d'entrainment diminue Lorsque $s \uparrow$ car moins de pénalisation permet un meilleur ajustement au donnée

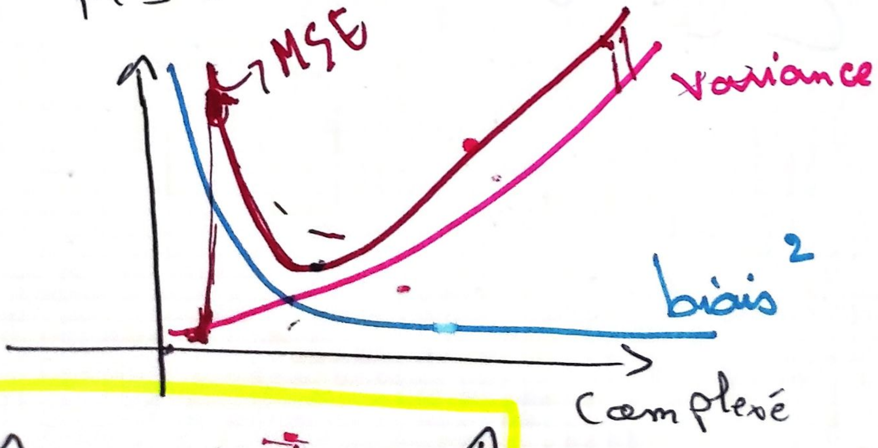


2 - Faux $\quad x = 0$

3 - F

4 - V

5 : MSE = $\underbrace{\text{biais}^2 + \text{variance}}$



$\left\{ \text{variance tjrs : croissante} \right.$

$\boxed{\lambda \uparrow \rightsquigarrow MSE \uparrow}$

5 - Faux

6 - Faux

7 - ~~Vrai~~ Faux

8 - ~~Faux~~ Vrai

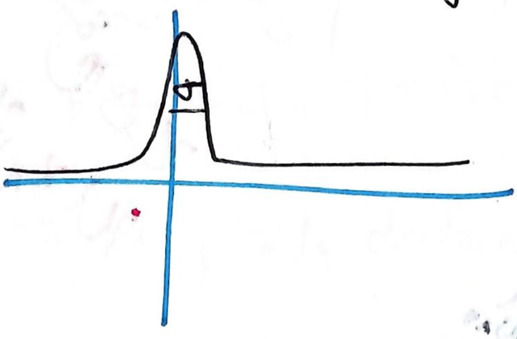
dans le cas globale variance

variance entre β

variance entre β va diminuer si on a une forte pénalité

$\lambda \uparrow \rightsquigarrow$ Variance $\searrow$; $\beta \searrow$

$\sigma \searrow$



Question 15 :

1 - partie differentiable et non :

* differentiable

$$\frac{1}{N} \sum \left(t^i - \beta_0 + \sum_{d=1}^{D} \beta_d x_d^i \right) + \lambda_2 \sum \beta_d^2$$

* non-differentiable :

$$\lambda_1 \sum_{d=1}^{D} |\beta_d|$$

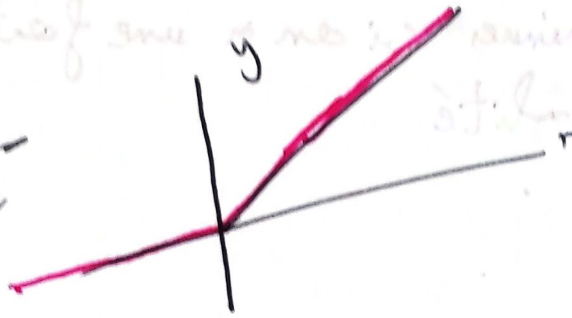
2) Voir figure 5

Lasso

Ridg

Question 7

1-
a

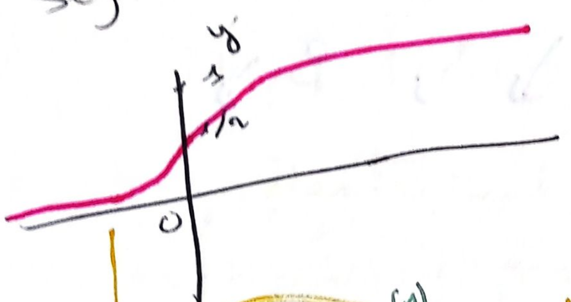


$$\begin{cases} y = 0 & \text{si } x \leq 0 \\ y > 0 & \text{sinon} \end{cases}$$

Relu

$$\text{Relu} = \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} = \frac{x + |x|}{2}$$

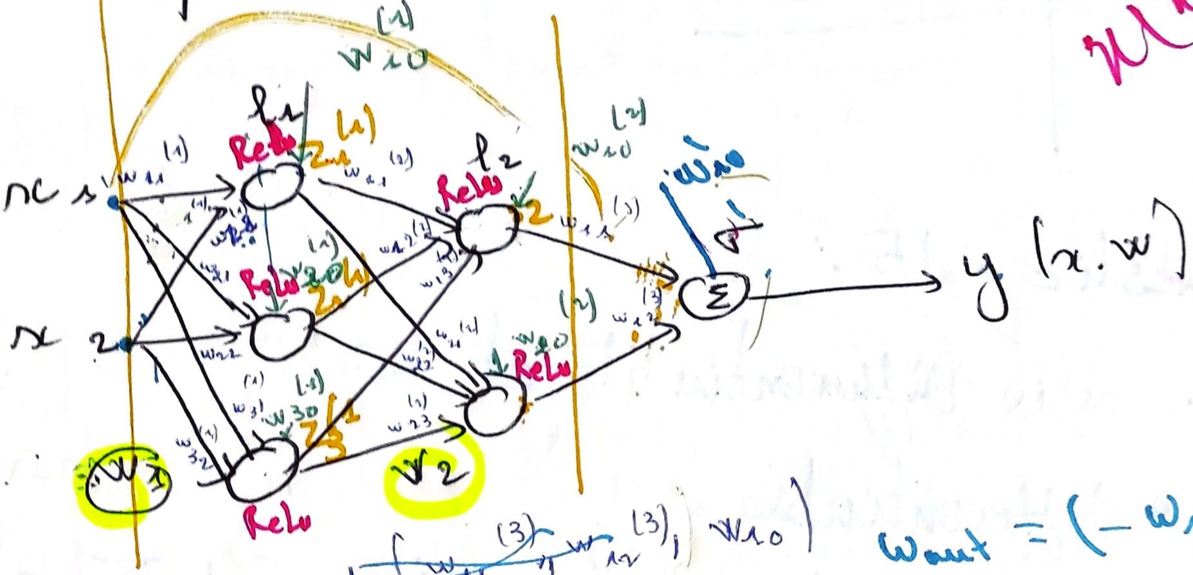
b - Segmoid



$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\sigma(x) = \begin{cases} 1 & x > 0 \\ 1/2 & x = 0 \\ 0 & \text{sinon} \end{cases}$$

2 -



~~w_{out}~~ $(w_{11}^{(3)}, w_{12}^{(3)}, w_{10})$ $\quad w_{out} = (-w_{10} -)$

$$w_{out}\left(w_{11}^{(3)}, w_{12}^{(3)}, w_{10}^{(3)}\right).$$

$$\boxed{y(x, w) = \sigma\left(\underbrace{\text{Relu}(X\, w_2^T + w_1)}_{\text{la sortie précedente}} \times w_{out}\right)}$$

poids actuelle contient w_0

1- Relu (w, x)

~~X~~ = Relu(w_2 . Relu $(w_1 x)$) $\Rightarrow y = \sigma(X \cdot w_{out})$

- couche 1 :

Relu ($w_1 \, x$)

- couche 2 :

Relu (w_2 Relu ($w_1 . x$))

→ Sortie

σ (Relu (w_2 Relu ($w_1 . x$)

$x = [1, x]$ car les matrice w_1 et w_2 contient les intercepts

explication :

- w_1, w_2 : contient intercepts
- w_1 poids de couche 1
- w_2 — — — — 2
- w_{out} poids de dernier sortie

d'où : $y(x, w_1, w_2) = \sigma\left(\text{Relu}\left(w_2 . \text{Relu}\left(w_1 . [1, x]^T\right)\right)\right)$

Question 8 :

least squares classifier : classificateurs des moindres carrés