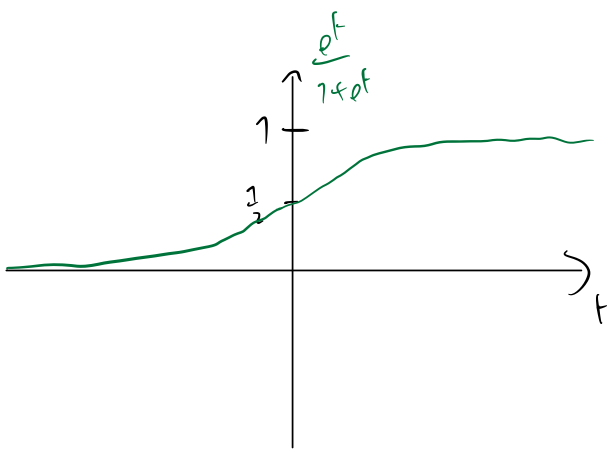
Exercice 1

1.) $\dfrac{e^0}{1+e^0} = \dfrac{1}{2}$ $\qquad \lim\limits_{t\to-\infty} \dfrac{e^t}{1+e^t} = \dfrac{0}{1} = 0$

$$\lim_{t\to+\infty} \frac{e^t}{1+e^t} = \lim_{t\to+\infty} \frac{e^t}{1+e^t} = 1$$

$$\frac{\partial}{\partial t}\left(\frac{e^t}{1+e^t}\right) = \frac{e^t(1+e^t) - e^t e^t}{(1+e^t)^2} = \frac{e^t}{(1+e^t)^2} \overset{(t=0)}{=} \frac{1}{2^2} = \frac{1}{4}$$



2)
$$\frac{e^{x}}{1+e^{x}} \geq \frac{1}{2}$$

$$\Leftrightarrow \quad e^{x} \geq \frac{1}{2} + \frac{1}{2} e^{x}$$

$$\Leftrightarrow \quad \frac{1}{2} e^{x} \geq \frac{1}{2}$$

$$\Leftrightarrow \quad e^{x} \geq 1$$

$$\Leftrightarrow \quad x \geq 0$$

3) $C_{\alpha,\beta}(x) \geq \frac{1}{2} \Leftrightarrow \frac{e^{x}}{1+e^{x}} \geq \frac{1}{2}$

donc $\begin{cases} x \geq 0 & \text{classé } 1 \\ x < 0 & \text{classé } 0 \end{cases}$

3) $C_{\alpha,\beta}(x) \geq \frac{1}{2} \Leftrightarrow \frac{e^{1-x}}{1+e^{1-x}} \geq \frac{1}{2}$

On a vu que en général $\frac{e^{t}}{1+e^{t}} \geq \frac{1}{2} \Leftrightarrow t \geq 0$

Donc $C_{\alpha,\beta}(x) \geq \frac{1}{2} \Leftrightarrow 1-x \geq 0$

$\Longleftrightarrow \quad x \leq 1$

Tout les points de $[-1, 1]$ sont cassès 1

4) Si $x > 0$ $\quad \lim\limits_{\beta \to \infty} \dfrac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} = 1$

donc x devient classé 1

Si $x < 0$ $\quad \lim\limits_{\beta \to -\infty} \dfrac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} = 0$

donc x devient classé 0

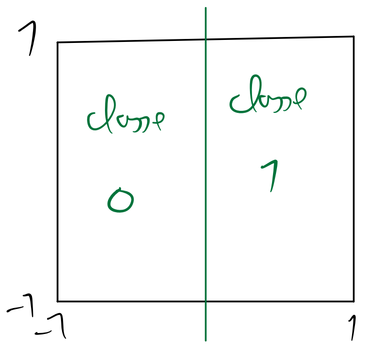
5) C'est le contraire (memes arguments)

$$\begin{cases} x > 0 & x \text{ classé } 0 \\ x < 0 & x \text{ classé } 1 \end{cases}$$

Exercice 2

1.) $C_{\alpha,\beta_1,\beta_2}(x) \geq \frac{1}{2} \iff \dfrac{e^{x_1}}{1+e^{x_1}} \geq \frac{1}{2}$

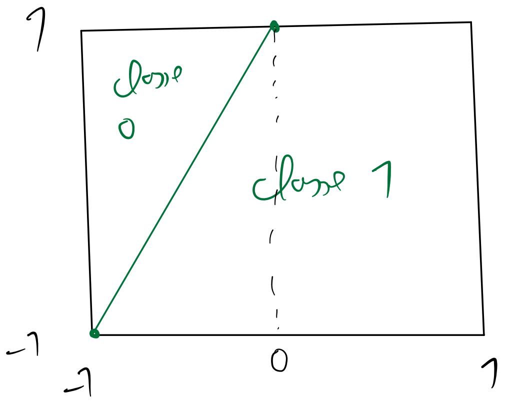
$\iff x_1 \geq 0$



2.) $C_{\alpha,\beta_1,\beta_2}(x) \geq \frac{1}{2} \iff \dfrac{e^{1+2x_1-x_2}}{1+e^{1+2x_1-x_2}} \geq \frac{1}{2}$

$\iff 1+2x_1-x_2 \geq 0$

$\iff x_2 \leq 1+2x_1$



3) $$C_{\alpha,\beta_1,\beta_2}(x) \geq \frac{1}{2} \iff \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2}}{1+e^{\alpha+\beta_1 x_1+\beta_2 x_2}} \geq \frac{1}{2}$$

$$\iff \alpha+\beta_1 x_1+\beta_2 x_2 \geq 0$$

Donc les x classés 1 sont $\{(x_1,x_2) ; \alpha+\beta_1 x_1+\beta_2 x_2 \geq 0\}$

La zone de séparation $\{(x_1,x_2) ; \alpha+\beta_1 x_1+\beta_2 x_2 = 0\}$

est une droite

Exercice 3

1) Pour x fixé

$$P(f(X) \neq Y \mid X = x)$$

$$= P(f(x) \neq Y \mid X = x)$$

Si $P(Y=1 \mid X=x) \geq \frac{1}{2}$

alors $c_\beta(x) \geq \frac{1}{2}$ et $f(x) = 1$

donc $P(f(x) \neq Y \mid X = x)$

$$= P(Y=0 \mid X=x) = 1 - P(Y=1 \mid X=x)$$

$$= 1 - c_\beta(x)$$

$$= \min(c_\beta(x), 1 - c_\beta(x))$$

Si $P(Y=1 \mid X=x) < \frac{1}{2}$ alors $c_\beta(x) < \frac{1}{2}$ et $f(x) = 0$

donc $P(f(x) \neq Y \mid X = x)$

$$= P(Y=1 \mid X=x)$$

$$= c_\beta(x)$$
$$= \min(c_\beta(x), 1 - c_\beta(x))$$

Donc $P(f(x) \neq Y \mid X = x) = \min(c_\beta(x), 1 - c_\beta(x))$

Donc $P(f(X) \neq Y)$

$$= E[\min(c_\beta(X), 1 - c_\beta(X))]$$

$$= \frac{1}{2^d} \int_{[-1,1]^d} \min(c_\beta(x), 1 - c_\beta(x))\, dx$$

3) Pour x fixé tel que $g(x) \neq f(x)$

$$P(Y \neq g(x) \mid X = x)$$
$$= P(Y = f(x) \mid X = x)$$

Si $c_\beta(x) \geq \frac{1}{2}$ alors $f(x) = 1$ et $P(Y = 1 \mid X = x) = c_\beta(x)$
$$= \max(c_\beta(x), 1 - c_\beta(x))$$

Si $\quad c_\beta(x) < \frac{1}{2}$ alors $f(x) = 0$ et

$$P(Y=0 \mid X=x) = 1 - c_\beta(x) = \max(c_\beta(x), 1 - c_\beta(x))$$

Donc $\quad P(Y \neq g(x) \mid X=x)$

$$= \mathbb{1}_{g(x)=f(x)} \quad P(Y \neq f(x) \mid X=x)$$

$$+ \mathbb{1}_{g(x)\neq f(x)} \max(c_\beta(x), 1 - c_\beta(x))$$

$$= \mathbb{1}_{g(x)=f(x)} \min(c_\beta(x), 1 - c_\beta(x)) \Big\} \text{ question précédente}$$

$$+ \mathbb{1}_{g(x)\neq f(x)} \max(c_\beta(x), 1 - c_\beta(x))$$

Donc

$$P(Y \neq g(X)) = \frac{1}{2^d} \int_{[-1,1]^d} \Big[\mathbb{1}_{g(x)=f(x)} \min(c_\beta(x), 1 - c_\beta(x))$$

$$+ \mathbb{1}_{g(x)\neq f(x)} \max(c_\beta(x), 1 - c_\beta(x)) \Big] dx$$

4) $\quad P(Y \neq g(X)) - P(Y \neq f(X))$

$$= \frac{1}{2^d} \int_{(-1,1)^d} \Big[\mathbb{1}_{g(x) = f(x)} \min(c_\beta(x), 1 - c_\beta(x))$$

$$+ \mathbb{1}_{g(x) \neq f(x)} \max(c_\beta(x), 1 - c_\beta(x)) \Big] dx \Bigg\} \text{ Q3}$$

$$- \frac{1}{2^d} \int_{(-1,1)^d} \Big[\mathbb{1}_{g(x) = f(x)} \min(c_\beta(x), 1 - c_\beta(x))$$

$$+ \mathbb{1}_{g(x) \neq f(x)} \min(c_\beta(x), 1 - c_\beta(x)) \Big] dx \Bigg\} \text{ Q1}$$

$$= \frac{1}{2^d} \int_{(-1,1)^d} \underbrace{\mathbb{1}_{g(x) \neq f(x)}}_{\geq 0} \Big[\underbrace{\max(c_\beta(x), 1 - c_\beta(x)) - \min(c_\beta(x), 1 - c_\beta(x))}_{\geq 0} \Big] dx$$

≥ 0

Exercice 4

1) $$P(Y_1=y_1, \dots, Y_n=y_n \mid X_1=x_1, \dots, X_n=x_n)$$

$$= P(Y_1=y_1 \mid X_1=x_1) \dots P(Y_n=y_n \mid X_n=x_n)$$

$$= \prod_{i=1}^{n} \left(\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right)^{y_i} \left(\frac{1}{1+e^{x_i^T\beta}}\right)^{1-y_i}$$

En effet $$\left(\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right)^{y_i} \left(\frac{1}{1+e^{x_i^T\beta}}\right)^{1-y_i} = \begin{cases} \dfrac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}} & \text{si } y_i=1 \\ 1-\dfrac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}} & \text{si } y_i=0 \end{cases}$$

2) $$-\log(P(\beta))$$

$$= -\log\left[\prod_{i=1}^{n} \left(\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right)^{y_i} \left(\frac{1}{1+e^{x_i^T\beta}}\right)^{1-y_i}\right]$$

$$= -\sum_{i=1}^{n}\left[y_i \log\left(\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right) + (1-y_i)\log\left(\frac{1}{1+e^{x_i^T\beta}}\right)\right]$$

$$= -\sum_{i=1}^{n}\Big[y_i x_i^T\beta - y_i \log(1+e^{x_i^T\beta})$$
$$- \log(1+e^{x_i^T\beta}) + y_i \log(1+e^{x_i^T\beta}) \Big]$$

$$= \sum_{i=1}^{n}\left[\log(1+e^{x_i^T\beta}) - y_i x_i^T\beta \right]$$

$$\frac{\partial}{\partial \beta_k} f(\beta) = \sum_{i=1}^{n}\left[\frac{e^{x_i^T\beta} x_{ik}}{1+e^{x_i^T\beta}} - y_i x_{ik} \right]$$

$$\frac{\partial}{\partial \beta_l}\frac{\partial}{\partial \beta_k} f(\beta) = \sum_{i=1}^{n} x_{ik}\left(\frac{x_{il}\, e^{x_i^T\beta}(1+e^{x_i^T\beta})}{(1+e^{x_i^T\beta})^2}\right.$$

$$\left. - \frac{e^{x_i^T\beta} x_{il}\, e^{x_i^T\beta}}{(1+e^{x_i^T\beta})^2} \right)$$

$$= \sum_{i=1}^{n} x_{ik}\, x_{il} \frac{e^{x_i^T \beta}}{\left(1+e^{x_i^T \beta}\right)^2}$$

$$M(\beta) = \sum_{i=1}^{n} \frac{e^{x_i^T \beta}}{\left(1+e^{x_i^T \beta}\right)^2} x_i x_i^T$$

3) $\forall v$ avec $\|v\| = 1$

$$v^T M(\beta) v = \sum_{i=1}^{n} \frac{e^{x_i^T \beta}}{\left(1+e^{x_i^T \beta}\right)^2} v^T x_i x_i^T v$$

$$= \sum_{i=1}^{n} \frac{e^{x_i^T \beta}}{\left(1+e^{x_i^T \beta}\right)^2} \left(v^T x_i\right)^2$$

$$\geq \left(\min_{i=1,\dots,n} \frac{e^{x_i^T \beta}}{\left(1+e^{x_i^T \beta}\right)^2}\right) \sum_{i=1}^{n} \left(v^T x_i\right)^2$$

$$= \left(\min_{i=1,\dots,n} \frac{e^{x_i^T \beta}}{\left(1+e^{x_i^T \beta}\right)^2}\right) \sum_{i=1}^{n} v^T x_i x_i^T v$$

$$= \left(\min_{i=1,\dots,n} \frac{e^{x_i^T \beta}}{(1+e^{x_i^T \beta})^2} \right) u^T \left(\sum_{i=1}^{n} x_i x_i^T \right) u$$

$$= \left(\min_{i=1,\dots,n} \frac{e^{x_i^T \beta}}{(1+e^{x_i^T \beta})^2} \right) u^T X^T X u$$

$$\geq c(\beta) \|u\|^2$$

4) Comme $g(\beta) \underset{\|\beta\| \to \infty}{\longrightarrow} +\infty$

Il existe $R < \infty$ tel que

$$\inf_{\|\beta\| \geq R} g(\beta) > g(0)$$

Sur le compact $B_R = \{ \beta ; \|\beta\| \leq R \}$ la fonction $g(\beta)$ est continue, donc elle admet au moins un minimiseur β^*

$$\forall \beta \in B_R \quad g(\beta) \geq g(\beta^*)$$

$$\forall \beta \text{ avec } \|\beta\| \geq R, \ g(\beta) > g(0) \geq g(\beta^*)$$

Donc β^* est un minimiseur global

Supposons qu'il existe $\bar{\beta}$ avec $f(\bar{\beta}) = f(\beta^*)$
et $\bar{\beta} \neq \beta^*$

On regarde la fonction $g : [0,1] \to \mathbb{R}$

définie par $g(t) = f\left[\beta^* + t(\bar{\beta} - \beta^*)\right]$

alors $g(0) = g(1)$

et pour tout $t \in [0,1]$ $g(t) \geq g(0)$

g est strictement convexe d'après Question 3
(cf cours optimisation)

Donc pour tout $t \in]0,1[$

$$g(t) = g((1-t)0 + t.1)$$

$$< (1-t)\, g(0) + t\, g(1)$$

$$= g(0)$$

ce qui est absurde.

Donc β^* est bien l'unique minimiseur
global

Exercice 3 question 2

$$P(C_\beta(X) \neq Y) = \frac{1}{2^d} \int_{[-1,1]^d} \min\left(C_\beta(x), 1 - C_\beta(x)\right) dx$$

On pose $f(\beta^T x) = \frac{1}{2^d} \min\left(\frac{e^{\beta^T x}}{1+e^{\beta^T x}}, 1 - \frac{e^{\beta^T x}}{1+e^{\beta^T x}} \right)$

alors $P(C_\beta(X) \neq Y) = \int_{\mathbb{R}^d} \mathbb{1}_{x \in [-1,1]^d} f(\beta^T x)\, dx$

$$= \int_{\mathbb{R}^d} \mathbb{1}_{x \in [-1,1]^d} f\left(\|\beta\| \tilde{\beta}^T x\right) dx$$

avec $\tilde{\beta} = \frac{\beta}{\|\beta\|}$

$$\leq \int_{\mathbb{R}^d} \mathbb{1}_{\|x\| \leq \sqrt{d}} f\left(\|\beta\| \tilde{\beta}^T x\right) dx$$

On va faire un changement de variables multi-dimensionnel.

La formule générale est :

Changement de variable en dimension d :

Soit $\psi : \mathbb{R}^d \to \mathbb{R}$

et φ bijective de $\mathbb{R}^d \to \mathbb{R}^d$

et $A \subset \mathbb{R}^d$

Alors

$$\int_A \psi \circ \varphi(x) \det\left[(J\varphi)(x)\right] dx = \int_{\varphi(A)} \psi(x)\, dx$$

avec $\varphi(A) = \{\varphi(x) ; x \in A\}$

$J\varphi(x)$ la matrice $d \times d$ avec $(J\varphi(x))_{ij} = \dfrac{\partial \varphi_i(x)}{\partial x_j}$

En fait ça généralise le changement de variable à une variable

$$\int_a^b \psi \circ \varphi(t)\, \varphi'(t)\, dt = \int_{\varphi(a)}^{\varphi(b)} \psi(t)\, dt$$

On applique le changement de variable multi-dimensionnel.

Pour A on prend $\mathbb{R}^d$

Pour $t = \begin{pmatrix} t_1 \\ \vdots \\ t_d \end{pmatrix} \in \mathbb{R}^d$ on prend $\psi(t) = \mathbb{1}_{\|t\| \leq \sqrt{d}} \, \beta(\|\beta\| t_1)$

On étend $\tilde{\beta}$ en une base orthonormale $\tilde{\beta}, e_2, \dots, e_d$.

et on pose
$$\varphi(t) = \begin{pmatrix} \tilde{\beta}^T t \\ e_2^T t \\ \vdots \\ e_d^T t \end{pmatrix}$$

Alors on vérifie que $(J\varphi)(t) = \begin{pmatrix} \tilde{\beta}^T \\ e_2^T \\ \vdots \\ e_d^T \end{pmatrix}$

et donc (matrice orthonormale) $\det\left((J\varphi)(t)\right) = 1$

Ainsi avec le changement de variables

$$\int_{\mathbb{R}^d} \mathbb{1}_{\|x\| \leq \sqrt{d}}\, \theta\left(\|\beta\|\, \tilde{\beta}^T x\right) dx$$

$$= \int_{\varphi(\mathbb{R}^d)} \underbrace{\mathbb{1}_{\|\varphi(x)\| \leq \sqrt{d}}\, \theta\left(\|\beta\|\, \varphi(x)_1\right)}_{\psi \circ \varphi(t)} \underbrace{1}_{\det((J\varphi)(t))} dx$$

car φ est bijective de $\mathbb{R}^d$ dans $\mathbb{R}^d$

et $\|\varphi(t)\| = \|t\|$ (matrice orthonormale)

$$= \int_{\mathbb{R}^d} \mathbb{1}_{\|t\| \leq \sqrt{d}}\, \theta\left(\|\beta\|\, t_1\right) dx$$

Lorsque $t_1 \neq 0$ alors $\theta(\|\beta\|\, t_1) \xrightarrow[\|\beta\| \to \infty]{} 0$

et $\left| \mathbb{1}_{\|t\| \leq \sqrt{d}}\, \theta(\|\beta\|\, t_1) \right| \leq \mathbb{1}_{\|t\| \leq \sqrt{d}}$

est intégrable

Donc d'après le théorème de convergence dominée (cf cours d'intégration)

$$P(C_\beta(x) \neq Y) \underset{\|\beta\| \to \infty}{\longrightarrow} 0$$

L'intuition de ce résultat est que lorsque $\|\beta\|$ est grand alors le problème de classification est plus facile, car pour la plupart des x, $\beta^T x$ est proche de $+\infty$ ou $-\infty$ donc $P(Y=1 \mid X=x)$ est proche de 0 ou 1 et donc Y est bien prédictible (peu d'aléa) lorsque x est connu.