Computational Statistics : Assignment 1

Raphaël Bernas

October 2024

1 Exercice 1

During all this work, h is a bounded and measurable function.

1.1 Q1.

Let us define $R \sim \mathcal{R}$ the Rayleigh distribution such that for all $r \in \mathrm{R}$ we have $f_R(r) = r\exp(\frac{-r^2}{2})\mathbf{1}_{\mathrm{R}^+}(r)$ and $\Theta \sim \mathcal{U}(0, 2\pi)$.
Assume that $X = R\cos(\Theta)$ and $Y = R\sin(\Theta)$, we want to prove that X and Y are independant with a centered-normed gaussian distribution.
Recall that $\mathbb{E}[h(X,Y)] = \int_{R^2} h(x,y) f_{(X,Y)}(x,y)dxdy$.

$$\mathbb{E}[h(R\cos(\Theta), R\sin(\Theta))] = \int_{R^2} h(r\cos(\theta), r\sin(\theta)) r\exp(\frac{-r^2}{2})\mathbf{1}_{\mathrm{R}^+}(r)\frac{1}{2\pi}drd\theta$$

$$|det(J_\theta)|^{-1} = \frac{1}{r}$$

where J_θ is the jacobian of our transformation $(X,Y) = \Phi(R,\Theta)$.

$$\mathbb{E}[h(R\cos(\Theta), R\sin(\Theta))] = \int_{R^2} h(x,y)\exp(\frac{-x^2-y^2}{2})\frac{1}{2\pi}dxdy$$

$$= \int_{R^2} h(x,y)\exp(\frac{-x^2}{2})\exp(\frac{-y^2}{2})\frac{1}{2\pi}dxdy$$

Which shows independance thanks to the fact that the density of both is separable into the product of the density of each.

1.2 Q2.

Let us compute :

$$\mathbb{P}(R \leq r) = \int_0^r x\exp(\frac{-x^2}{2})dx$$

$$= [-\exp(\frac{-x^2}{2}]_0^r$$

Algorithm 1: Gaussian distribution sampling algorithm

Sample $U \sim \mathcal{U}(0,1)$ and $U_\theta \sim \mathcal{U}(0,2\pi)$
$R = F^{-1}(U_\theta)$
Return $(R\cos(U_\theta), R\sin(U_\theta))$

$$= 1 - \exp(\frac{-r^2}{2})$$

Assume we have $U \sim \mathcal{U}(0,1)$ then $F^{-1}(U) \sim \mathcal{R}$. Indeed $\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = \mathbb{P}(R \leq x)$. Thus we define Algorithm 1 which permit us to generate gaussian samples.

1.3 Q3.

a) (V_1, V_2) follow the uniform distribution on the unite ball in $\mathbb{R}^2$.
b) T is the number of loop, here it can be seen as a random variable following a geometric law with a parameter $\frac{\pi}{4}$. Indeed, it is the surface area of the unite ball in $\mathbb{R}^2$ divided by the area of a square with side length equal to 2. Thus $\mathbb{P}(T = k) = P_{rej}^{k-1} P_{acc} = (1 - \frac{\pi}{4})^{k-1} \frac{\pi}{4}$.

$$\mathbb{E}(T) = \frac{1}{P_{acc}} = \frac{4}{\pi}$$

c) We will prove such a result the same way we have done in 1.1. This give us that

$$\mathbb{E}[h(T_1, V)] = \mathbb{E}[h(\frac{V_1}{\sqrt{V_1^2 + V_2^2}}, V_1^2 + V_2^2)]$$

We denote $\mathcal{D} = \{(x,y) \in [-1,1]^2 | x^2 + y^2 < 1\}$

$$= \int_{\mathcal{D}} h(\frac{v_1}{\sqrt{v_1^2 + v_2^2}}, v_1^2 + v_2^2) \frac{1}{\pi} dv_1 dv_2$$

$$= \int_{\mathcal{D} \cap \mathbb{R} \times \mathbb{R}^+} h(\frac{v_1}{\sqrt{v_1^2 + v_2^2}}, v_1^2 + v_2^2) \frac{1}{\pi} dv_1 dv_2 + \int_{\mathcal{D} \cap \mathbb{R} \times \mathbb{R}^-} h(\frac{v_1}{\sqrt{v_1^2 + v_2^2}}, v_1^2 + v_2^2) \frac{1}{\pi} dv_1 dv_2$$

We divided the integral so that the *cos* restricted on both set is a diffeomorphism. On one set we compute over $[0, \pi]$ and $[\pi, 2\pi]$.
However we have that the jacobian of our variable substitution is :

$$|det \begin{pmatrix} \frac{\sqrt{v_1^2 + v_2^2} - \frac{v_1^2}{\sqrt{v_1^2 + v_2^2}}}{v_1^2 + v_2^2} & \frac{-v_2^2}{(v_1^2 + v_2^2)^{\frac{3}{2}}} \\ 2v_1 & 2v_2 \end{pmatrix}|^{-1} = |\frac{2v_2}{\sqrt{v_1^2 + v_2^2}}|^{-1} = \frac{\sqrt{v_1^2 + v_2^2}}{2v_2}$$

Thus,

$$\mathbb{E}[h(T_1, V)] = \int_{[-1,1] \times [0,1]} h(t_1, v) \frac{1}{2\pi\sqrt{1 - t_1^2}} dt_1 dv + \int_{[-1,1] \times [0,1]} h(t_1, v) \frac{1}{2\pi\sqrt{1 - t_1^2}} dt_1 dv$$

$$= \int_{[-1,1]\times[0,1]} h(t_1, v) \frac{1}{\pi\sqrt{1-t_1^2}} dt_1 dv$$

Which imply that the density of both is equal to the product of both density, hence the expected independance result. Furthemore, we obtain that V follow the uniform law on $[0,1]$.
Now we want to find the law of T_1 :
First, let us note $\Theta \sim \mathcal{U}(0, 2\pi)$

$$\mathbb{E}(h(cos(\Theta)) = \int_{\mathbb{R}^2} h(cos(\theta)) \mathbf{1}_{[0,2\pi]}(\theta) \frac{1}{2\pi} d\theta$$

$$= \int_0^{\pi} h(cos(\theta)) \frac{1}{2\pi} d\theta + \int_{\pi}^{2\pi} h(cos(\theta)) \frac{1}{2\pi} d\theta$$

We compute the jabobian of the substitution $t = cos(\theta)$[1] : $\frac{1}{|sin(\theta)|} = \frac{1}{\sqrt{1-t^2}}$.
Thus :

$$\int_{\mathbb{R}^2} h(cos(\theta)) \mathbf{1}_{[0,2\pi]}(\theta) \frac{1}{2\pi} d\theta$$

$$= \int_{-1}^{1} h(t) \frac{1}{2\pi\sqrt{1-t^2}} dt + \int_{1}^{-1} h(t) \frac{1}{2\pi\sqrt{1-t^2}} d(-t)$$

$$= \int_{-1}^{1} h(t) \frac{1}{\pi\sqrt{1-t^2}} dt$$

Thus, $cos(\Theta)$ follow the same law as T_1.
d) Let us determine the law of (X, Y).

$$\mathbb{E}[h(X,Y)] = \int_{\mathbb{R}^2} h(\sqrt{-2log(v_1^2+v_2^2)} \frac{v_1}{\sqrt{v_1^2+v_2^2}}, \sqrt{-2log(v_1^2+v_2^2)} \frac{v_2}{\sqrt{v_1^2+v_2^2}}) \mathbf{1}_{\mathcal{D}}(v_1, v_2) dv_1 dv_2$$

$$= \int_{\mathbb{R}^2} h(\sqrt{-2log(v_1^2+v_2^2)} \frac{v_1}{\sqrt{v_1^2+v_2^2}}, \sqrt{-2log(v_1^2+v_2^2)} \frac{v_2}{\sqrt{v_1^2+v_2^2}}) \mathbf{1}_{\mathcal{D}}(v_1, v_2) dv_1 dv_2$$

Just as we have done for T_1, we can show that T_2 follow the law of $sin(\Theta)$. Thus let us rewrite the integral :

$$= \int_{[0,2\pi]\times[0,1]} h(\sqrt{-2log(v)} cos(\theta), \sqrt{-2log(v)} sin(\theta)) d\theta dv$$

We compute the jacobian of our substitution :

$$|det \begin{pmatrix} \frac{-1}{v\sqrt{-2log(v)}} cos(\theta) & -\sqrt{-2log(v)} sin(\theta) \\ \frac{-1}{v\sqrt{-2log(v)}} sin(\theta) & \sqrt{-2log(v)} cos(\theta) \end{pmatrix} |^{-1} = v = exp(-\frac{x^2+y^2}{2})$$

Thus :

$$= \int_{\mathbb{R}^2} h(x,y) exp(-\frac{x^2+y^2}{2}) dx dy$$

This implies that $(X, Y) \sim \mathcal{N}(0_2, Id_2)$.

[1] On both set, cos is a diffeomorphism.

2 Exercice 2

2.1 Q1.

Let us define the Markov chain $(X_n)_{n\geq 0}$ as follow :
- If $X_n = \frac{1}{k}$, then

$$\begin{cases} X_{n+1} = \frac{1}{k+1} & \text{with probability } 1 - X_n^2 \\ X_{n+1} \sim \mathcal{U}(0,1) & \text{with probability } X_n^2 \end{cases}$$

- If not, $X_{n+1} \sim \mathcal{U}(0,1)$.
As we have, for all A an element of the σ-algebra :

$$P(x,A) = \mathcal{P}(X_{n+1} \in A | X_n = x)$$

$$= \begin{cases} \mathbb{P}(X_{n+1} = \frac{1}{k+1})\mathbb{P}(X_{n+1} \in A | X_{n+1} = \frac{1}{k+1}) + \mathbb{P}(X_{n+1} \neq \frac{1}{k+1})\mathbb{P}(X_{n+1} \in A | X_{n+1} \neq \frac{1}{k+1}) & \text{if } x = \frac{1}{k} \\ \mathbb{P}(X_{n+1} \in A | X_{n+1} \sim \mathcal{U}(0,1)) & \text{otherwise} \end{cases}$$

$$= \begin{cases} (1-x^2)\delta_{\frac{1}{k+1}}(A) + x^2 \int_{A\cap[0,1]} dt & \text{if } x = \frac{1}{k} \\ \int_{A\cap[0,1]} dt & \text{otherwise} \end{cases}$$

2.2 Q2.

We want to prove that $\pi P = \pi$
Let us compute for all A in the σ-algebra :

$$\pi P(A) = \int_{\mathbb{R}} \pi(x) P(x,A) dx$$

As π is the uniform distribution on $[0,1]$, thus $\pi(B) = 0$ if $B \cap [0,1] = \emptyset$

$$= \int_{[0,1]} P(x,A) dx$$

Now we denote $\mathcal{K} = \{\frac{1}{k} \mid k \in \mathbb{N}^*\}$, we have :

$$\pi P(A) = \int_{[0,1]\cap\mathcal{K}} P(x,A) dx + \int_{[0,1]\cap\mathcal{K}^C} P(x,A) dx$$

However, $\int_{\mathcal{K}} dx = 0$ as $\mathcal{K}$ is a countable set. Thus, we get that computing a measurable function[2] upon such a set return value 0.

$$\pi P(A) = \int_{[0,1]} \int_{A\cap[0,1]} dt dx = \int_{A\cap[0,1]} dt = \pi(A)$$

[2]Indeed, $x \mapsto P(x,A)$ is measurable for all A because it is constant almost everywhere (the set on which it is not is of measure equal to 0 and the function on this set is bounded).

2.3 Q3.

Let $x \in \mathbb{R}$ and f a bounded measurable function. We want to compute

$$\begin{aligned} Pf(x) &= \mathbb{E}[f(X_1) \mid X_0 = x] \\ &= \int_{\mathbb{R}} P(x, dy) f(y) \\ &= \int_{[0,1]} P(x, dy) f(y) \\ &= \begin{cases} (1-x^2) f\left(\frac{1}{k+1}\right) + x^2 \int_0^1 f(y)\, dy & \text{if } x = \frac{1}{k} \\ \int_0^1 f(y)\, dy & \text{otherwise} \end{cases} \end{aligned}$$

Thus, for $n \geq 1$:

$$P^n f(x) = \begin{cases} (1-x^2) P^{n-1} f\left(\frac{1}{k+1}\right) + x^2 \int_0^1 P^{n-1} f(y)\, dy & \text{if } x = \frac{1}{k} \\ \int_0^1 P^{n-1} f(y)\, dy & \text{otherwise} \end{cases}$$

We note[3] $I_f = \int_0^1 f(y)\, dy$
Moreover we can show by induction that :

$$\int_0^1 P^n f(y)\, dy = \int_0^1 f(y)\, dy = I_f$$

This is obviously the case for $n = 0$ and suppose it is for n, then :

$$\begin{aligned} \int_0^1 P^{n+1} f(y)\, dy &= \int_{[0,1] \cap \mathcal{K}} P^{n+1} f(y)\, dy + \int_{[0,1] \cap \mathcal{K}^C} P^{n+1} f(y)\, dy \\ &= \int_{[0,1] \cap \mathcal{K}^C} P^{n+1} f(y)\, dy \\ &= \int_0^1 \left(\int_{[0,1]} P^n f(t)\, dt \right) dy \\ &= \int_0^1 I_f\, dy = I_f \end{aligned}$$

This give us :

$$P^n f(x) = \begin{cases} (1-x^2) P^{n-1} f\left(\frac{1}{k+1}\right) + x^2 I_f & \text{if } x = \frac{1}{k} \\ I_f & \text{otherwise} \end{cases}$$

Thus we get that for all $x \notin \mathcal{K}$ with $\mathcal{K}$ defined in 2.2 and for all $n \in \mathbb{N}^*$:

$$\begin{aligned} P^n f(x) &= I_f = \pi f \\ \lim_{n \to +\infty} P^n f(x) &= I_f \end{aligned}$$

[3]Note that I_f can also be writen as $\int_{\mathbb{R}} f(y)\pi(y)dy$.

2.4 Q4.

Let $x = \frac{1}{k}$ with $k \geq 2$.
a) We want to compute $P^n(x, \frac{1}{n+k}) = \mathbb{P}(X_n = \frac{1}{n+k} | X_0 = x)$. To do so first we need to remark that $\mathbb{P}(X_{n+1} \in \mathcal{K} | X_n \notin \mathcal{K}) = 0$.
Indeed, $\mathbb{P}(X_{n+1} \in \mathcal{K} | X_n \notin \mathcal{K}) \leq \mathbb{P}(X_{n+1} \in \mathcal{K} | X_{n+1} \sim \mathcal{U}(0,1)) = 0$. Thus the only way to get $X_n = \frac{1}{n+k}$, while $X_0 = x$ is to have the following sequences $(X_p = \frac{1}{p+k})_{p \in [|0,n|]}$.
Therefore :

$$P^n(x, \frac{1}{n+k}) = \mathbb{P}(X_n = \frac{1}{n+k} | X_{n-1} = \frac{1}{n-1+k}) \cdots \mathbb{P}(X_1 = \frac{1}{k+1} | X_0 = \frac{1}{k} = x)$$

$$= (1 - \frac{1}{(n-1+k)^2}) \cdots (1 - \frac{1}{k^2})$$

$$= (\frac{(n+k)(n-2+k)}{(n-1+k)^2}) \cdots (\frac{(k+1)(k-1)}{k^2})$$

$$= (\frac{n+k}{n-1+k})(\frac{k-1}{k})$$

b) Using the precedent result we have that :

$$\lim_{n \to +\infty} P^n(x, \frac{1}{n+k}) = \frac{k-1}{k} > 0$$

Thus, for $A = \bigcup_{q \in \mathbb{N}} \left\{ \frac{1}{k+q+1} \right\}$:

$$\lim_{n \to +\infty} P^n(x, A) \geq \lim_{n \to +\infty} P^n(x, \frac{1}{n+k}) > 0$$

And as $\pi(A) = 0$ we have that $\lim_{n \to +\infty} P^n(x, A) \neq \pi(A)$.

3 Exercice 3

Let us define $\{z_i = (x_i, y_i) | i \in [|1, n|]\}$ our samples set. We want to solve :

$$\inf_{w \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i)^2 \tag{1}$$

We denote $J(w, (x_i, y_i)) = (y_i - w^T x_i)^2$. Thus (1) is equivalent to :

$$\inf_{w \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} J(w, (x_i, y_i)) \tag{2}$$

Finally, we will note $J_n(w) = \frac{1}{n} \sum_{i=1}^{n} J(w, (x_i, y_i))$.
Remark : $J(\cdot, (x_i, y_i))$ is a differenctiable function. Thus $J_n(\cdot)$ too.

Algorithm 2: Stochastic gradient descent

Input: Initial parameters w^0, learning rate $(\eta_k)_{k\in\mathbb{N}}$, batch lenght N_{batch}, precision ϵ, samples (x_i, y_i).

Initialize :

$w_0 \leftarrow w^0$

$k \leftarrow 1$

while $\frac{J_n(w^k)-J_n(w^{k-1})}{J_n(w^{k-1})} > \epsilon$ **do**

  $S \leftarrow 0$

  for $i = 1$ *to* N_{batch} **do**

    Sample $I \sim \mathcal{U}([|1, n|])$

    $S \leftarrow S - 2x_I(y_I - (w^k)^T x_I)$

  $S \leftarrow \frac{1}{N_{batch}} S$

  $w^{k+1} \leftarrow w^k - \eta_k S$

  $k \leftarrow k + 1$

Return w^k

3.1 Q1.

First we compute :

$$\nabla_w J(w, (x_i, y_i)) = -2x_i(y_i - w^T x_i)$$

Therefore :

$$\nabla_w J_n(w) = \frac{1}{n}\sum_{i=1}^{n} \nabla_w J(w, (x_i, y_i))$$

$$= \frac{1}{n}\sum_{i=1}^{n} -2x_i(y_i - w^T x_i)$$

We define the stochastic gradient descent method (also called SGDM) in Algorithm 2.

3.2 Q2.

The code for this assignment can be found on this link (in file named TP1.py) :

$$https://github.com/Raphael-Bernas/$$

See Figure 1

3.3 Q3.

We compute Algorithm 2 for our samples and obtain $\hat{w} \approx 1.27$ which is near from the expected result : 1.3.

See Figure 2

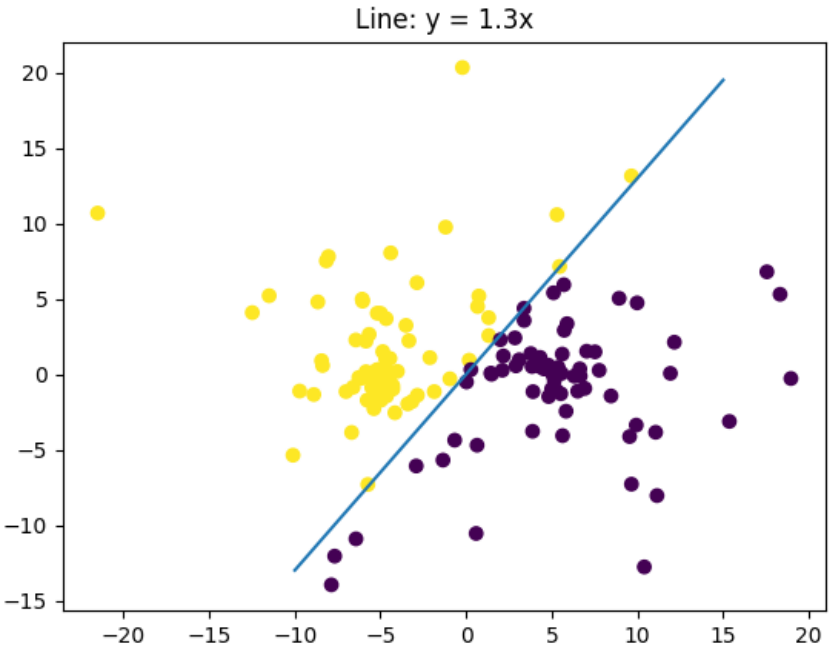

Figure 1: Q2. Classified sample divided by a line.

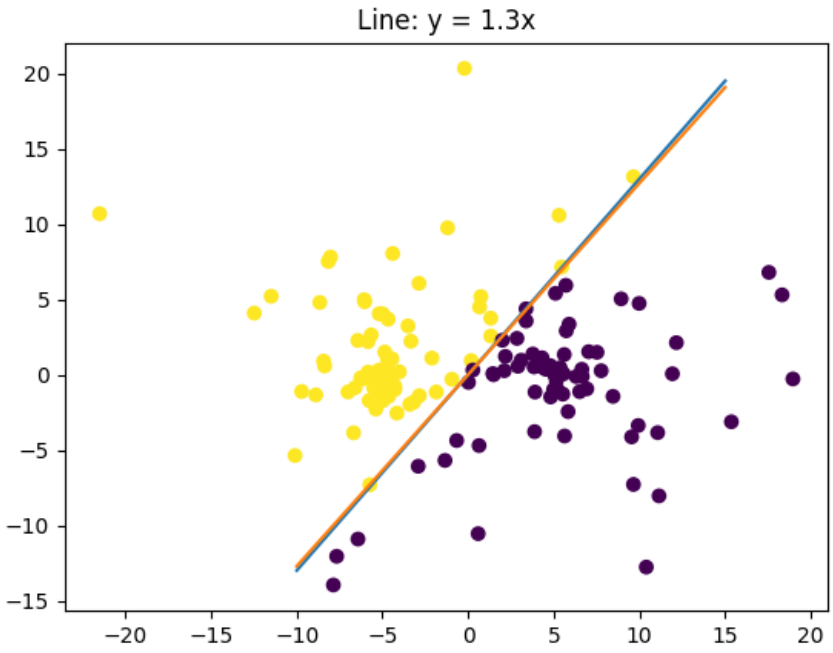

Figure 2: Q3. Classified sample divided by a line and the estimated line ($\hat{w} \approx 1.27$).

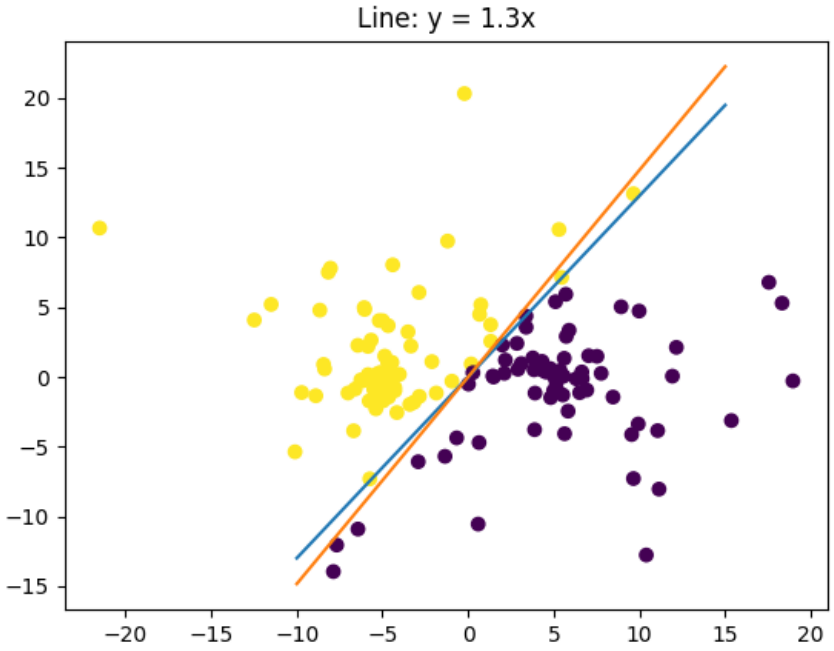

Figure 3: Q4. Classified sample divided by a line and an estimated line over noised data ($\hat{w}^{noised} \approx 1.48$).

3.4 Q4.

Now, we use the Marsaglia-Bray algorithm to generate a random gaussian noise that we add to our samples $(z_i)_i$:

$$z_i^{noised} = ((x_{1i}^{noised}, x_{2i}^{noised}), y_i)$$

$$= ((x_{1i} + \sigma\mathcal{N}(0,1), x_{2i} + \sigma\mathcal{N}(0,1), y_i)$$

We obtain $\hat{w}^{noised} \approx 1.48$ which is farther from the expected results than the value obtained without noise. For a visualization of such a result :

See Figure 3

3.5 Q5.

We compute a classification hyperplan using the stochastic gradient descent on the cancer dataset. However such a dataset has for each individual many

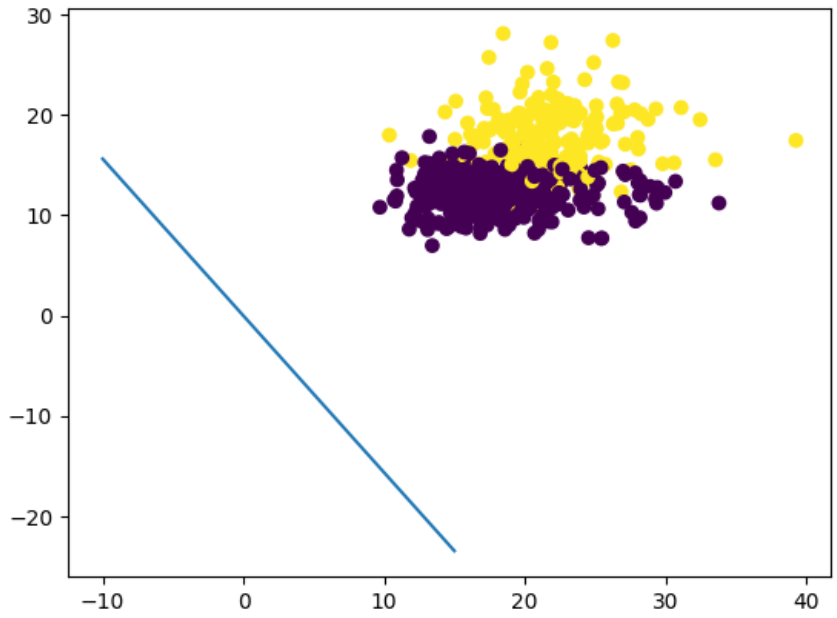

Figure 4: Q5. Cancer data set fitted with an hyperplan projected on the variable $\{radius1, texture1\}$: Projection on which the hyperplan model is **not accurate**.

features. Thus computing an hyperplan over all features space[4] can appear to be irrevelant. Figure 4 shows that in such a case computing a hyperplan is not accurate for such a complicated problem. However, there exist variable for which our hyperplan model appears to be working : See Figure 5. The main problem here is that for some variable, our dataset is not centered and scaled thus it leads to inproper reduction of the loss.

Therefore we decide to normalize the data with the following :

$$\tilde{X} = \frac{X - \mathbb{E}(X|\mathbb{P}_X)}{\sqrt{\mathbb{V}(X|\mathbb{P}_X)}}$$

Now we compute the loss function at each step with a learning rate decay $0 < \mu < 1$[5].

See Figure 6

[4] We mean here that each $x_i \in \mathbb{R}^d$, thus $w \in \mathbb{R}^d$.

[5] It means here that we compute $w^{k+1} \leftarrow w^k - (\mu)^k \eta \nabla_w J_n(w^k)$.

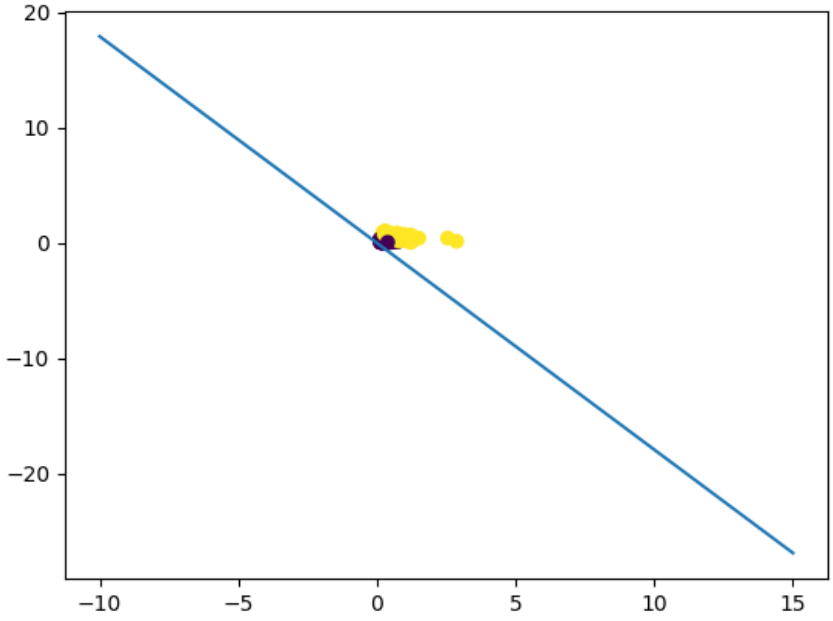

Figure 5: Q5. Cancer data set fitted with an hyperplan projected on the variable {*radius2,compactness3*} : Projection on which the hyperplan model is **accurate**.

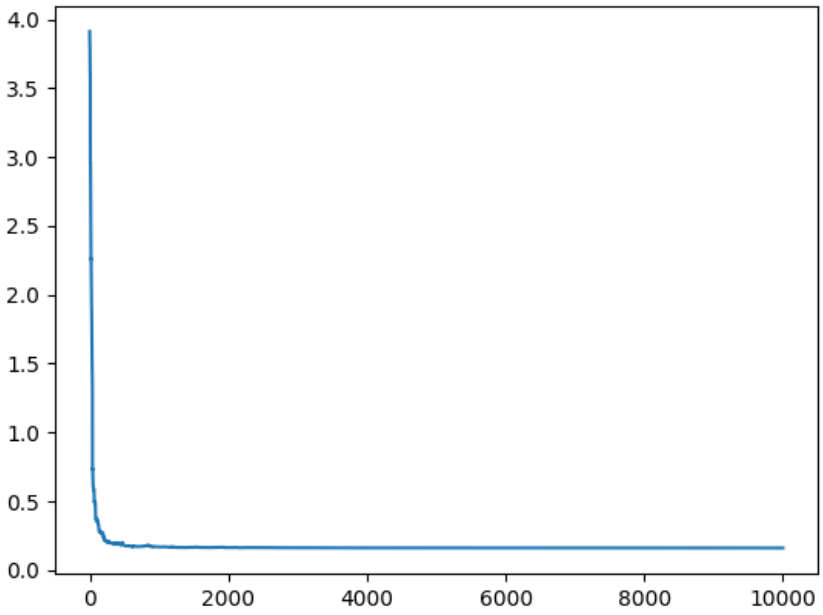

Figure 6: Q5. Cancer data set fitted with an hyperplan : Loss function evolution.