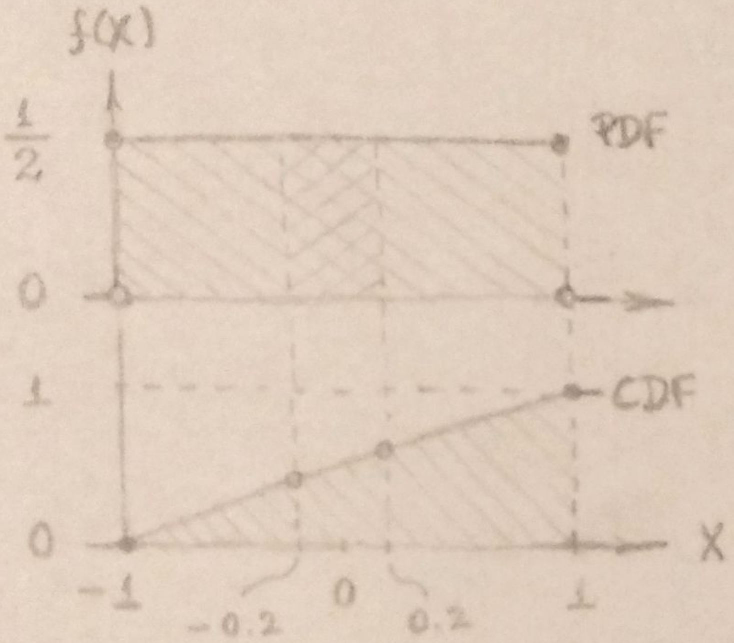
Lista 1

Aprendizado de máquina

Thiago Tomás de Paula

1)




a) The pdf requires that the probability of $X$ being in $(-1,1)$ be 100%. Therefore,

$$\int_{-1}^{1} pdf(x)\,dx = 1 \Rightarrow \int_{-1}^{1} p \cdot dx = 1 \therefore p = 1/2,$$

$p > 0$ constant by hypothesis.

b) Picture on the left.

$$cdf(x) = \int_{-1}^{x} pdf(x)\,dx = \frac{1}{2}(x+1)$$

c) $$P(-0.2 < X < 0.2) = \int_{-0.2}^{0.2} 1/2\,dx = 0.2 \,/\!/$$

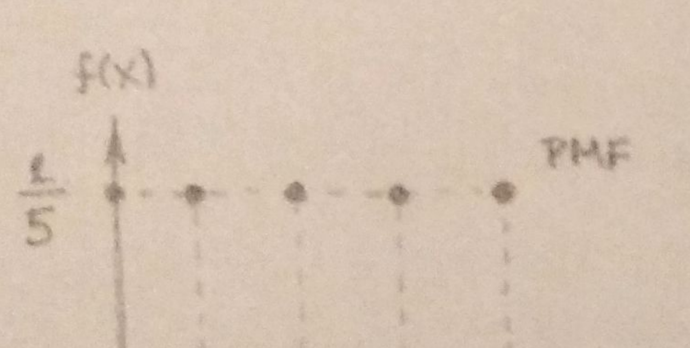
d) By definition,

$$E[X^n] = \int_{-1}^{1} x^n\, pdf(x)\,dx, \quad Var[X] = E[(X-E[X])^2] = E[X^2] - E[X]^2.$$

As such, we have

- $E[X] = \int_{-1}^{1} x \cdot \frac{1}{2}\,dx = \frac{1}{2} \cdot \frac{x^2}{2}\Big|_{-1}^{1} = \frac{1}{2}\cdot\left(\frac{1}{2} - \frac{1}{2}\right) = 0.$
- $E[X^2] = \int_{-1}^{1} x^2 \frac{1}{2}\,dx = \frac{1}{2}\cdot\frac{x^3}{3}\Big|_{-1}^{1} = \frac{1}{2}\cdot\left(\frac{1}{3} - \frac{-1}{3}\right) = \frac{1}{3}.$
- $Var[X] = \frac{1}{3} - 0 = \frac{1}{3}.$
- $E[X^4] = \int_{-1}^{1} x^4 \cdot \frac{1}{2}\,dx = \frac{1}{2}\cdot\frac{x^5}{5}\Big|_{-1}^{1} = \frac{1}{2}\cdot\left(\frac{1}{5} - \frac{-1}{5}\right) = \frac{1}{5}.$

2)




$$PMF = \begin{cases} p > 0, & X \in \{-2,-1,0,1,2\} := S \\ 0, & \text{otherwise} \end{cases}$$

$$\Rightarrow \sum_{x\in S} p = 1 \Leftrightarrow p = 1/5.$$

By definition,

$$E[X^n] = \sum_{x\in S} x^n\, PMF(x), \quad Var[X] = E[X^2] - E[X]^2$$

and so

- $E[X] = \sum_{x\in S} x \cdot \frac{1}{5} = \frac{1}{5}(-2-1+0+1+2) = 0$
- $E[X^2] = \sum_{x\in S} x^2 \frac{1}{5} = \frac{1}{5}(4+1+0+1+4) = 2$
- $Var[X] = E[X^2] - E[X] = 2 - 0 = 2.$

3) A gaussian joint pdf has general form

$$PDF(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{\frac{1}{2}}} \cdot \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$$

with

- $x = \begin{bmatrix} X_1 \\ \vdots \\ X_d \end{bmatrix}$, $X_n \sim N(\mu_n, \sigma_n^2)$, $1 \leq n \leq d$;
- $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix}$;
- $\Sigma_{ij} = \begin{cases} \sigma_i^2, & i=j \\ \sigma_{ij}, & i \neq j \end{cases}$, $1 \leq i \leq d$, $1 \leq j \leq d$.

In the question,

- $d = 2$;
- $\mu_1 = -2$, $\mu_2 = 1$;
- $\sigma_1^2 = 2$, $\sigma_2^2 = 4$;
- $\sigma_{12} = \sigma_{21} = -0.8$.

As such, we have

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \Rightarrow |\Sigma|^{d/2} = |\Sigma| = \sigma_1^2\sigma_2^2 - \sigma_{12}\cdot\sigma_{21} = 8 - 0.64 = 7.36,$$

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{21} & \sigma_1^2 \end{bmatrix} = \frac{1}{7.36}\begin{bmatrix} 4 & 0.8 \\ 0.8 & 2 \end{bmatrix},$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

This leads to

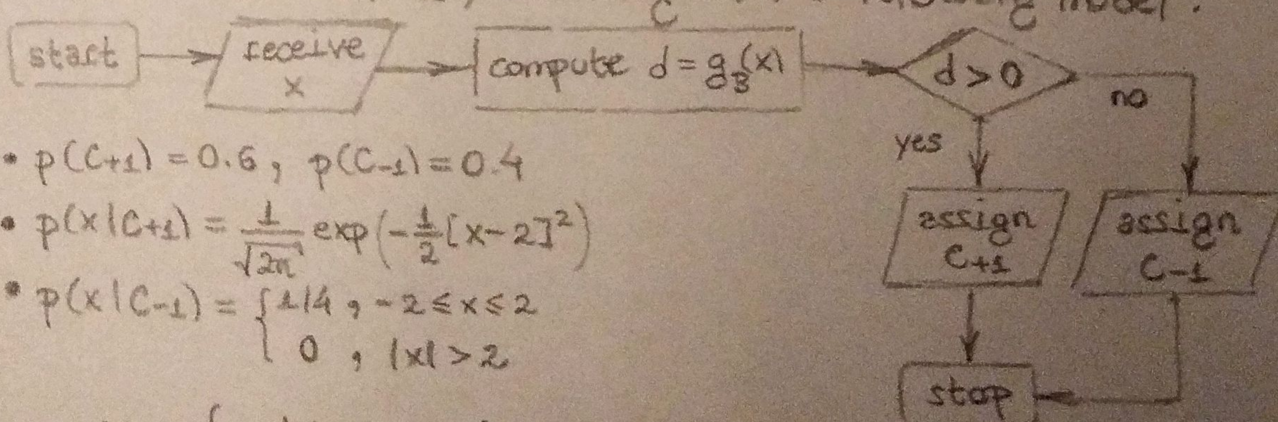
$$PDF(x) = \frac{1}{2\pi\sqrt{7.36}} \cdot \exp\left\{-\frac{1}{2}[x_1+2 \quad x_2-1]\frac{1}{7.36}\begin{bmatrix} 4 & 0.8 \\ 0.8 & 2 \end{bmatrix}\begin{bmatrix} x_1+2 \\ x_2-1 \end{bmatrix}\right\}.$$

4) a) Bayes' classifier uses Bayes' rule to classify an input, i.e.,
"Class assigned to $x$ maximizes $P(C_i|x) = \frac{p(x|C_i)p(C_i)}{p(x)}$".
Since $p(x)$ is class independent and we deal with only two (univariate) classes, the statement is equivalent to

"Choose $C_{+1}$ if $p(x|C_{+1})p(C_{+1}) > p(x|C_{-1})p(C_{-1})$ else $C_{-1}$".
Since $p(x|C_{+1})$ is gaussian and $p(x|C_{-1})$ piecewise constant, it makes sense to choose

$$g_{C_B}(x) = \log_e[p(x|C_{+1})] + \log_e[p(C_{+1})] - \log_e[p(x|C_{-1})] - \log_e[p(C_{-1})]$$

as the discriminant, resulting in the following model:




- $p(C_{+1}) = 0.6$, $p(C_{-1}) = 0.4$
- $p(x|C_{+1}) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}[x-2]^2\right)$
- $p(x|C_{-1}) = \begin{cases} 1/4, & -2 \leq x \leq 2 \\ 0, & |x| > 2 \end{cases}$

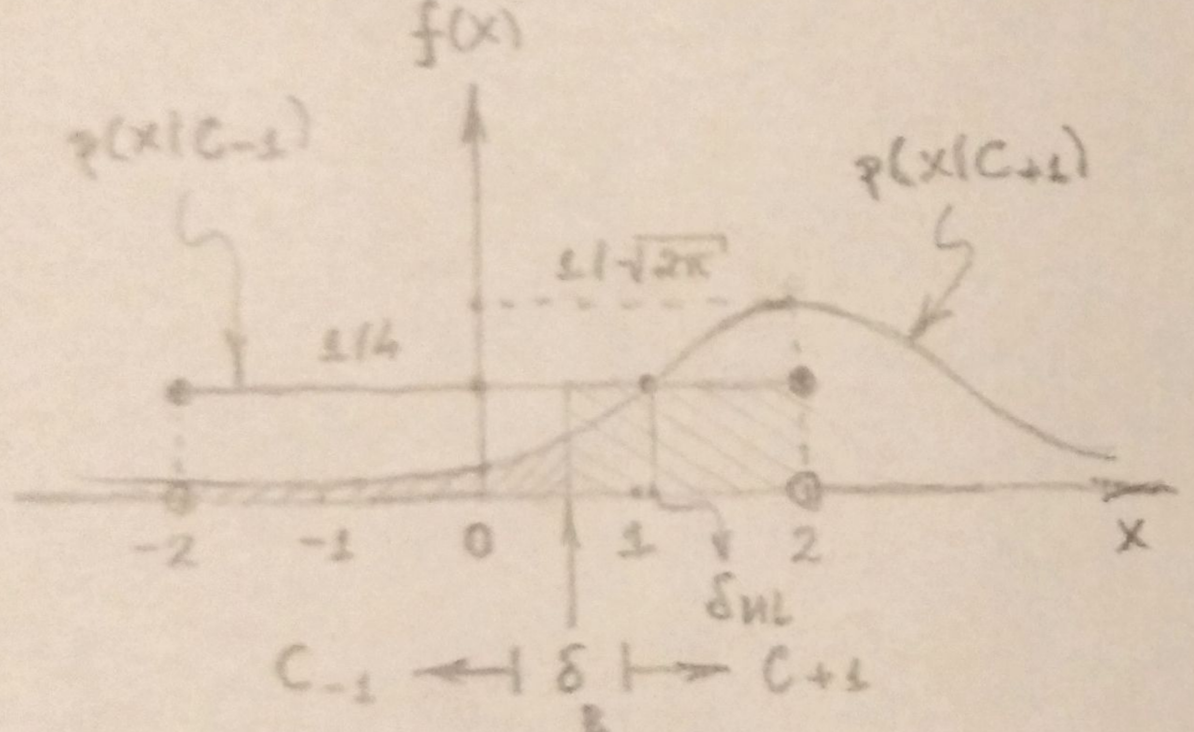
$$\Rightarrow g_{C_B}(x) = \begin{cases} -\frac{1}{2}\log_e(2\pi) - \frac{1}{2}(x-2)^2 + \log_e\frac{3}{2} - \log_e 1/4, & |x| \leq 2 \\ +\infty, & |x| > 2 \text{ (since } C_{-1} \text{ cannot happen)} \end{cases}$$

4) b) Boundary. The boundary is the set of x values such that the discriminant is null. We have

$$g_B(x) = 0 \iff -\frac{1}{2}\log 2\pi - \frac{1}{2}(x-2)^2 + \log\frac{3}{2} - \log\frac{1}{4} = 0 \quad (|x| \leq 2)$$

$$\iff |x-2| = \sqrt{\log\left(\frac{36}{2\pi}\right)} \quad \therefore \; x = 2 - \sqrt{\log\left(\frac{36}{2\pi}\right)} \approx 0.679 \; //$$

Error probability. The Bayes' method has set a boundary at $\delta = 2 - \sqrt{\log \frac{36}{2\pi}}$. From the drawing of the conditional PDFs below, it's clear that the method assigns $C_{+1}$ to $x \geq \delta$ and $C_{-1}$ to $x < \delta$.




The probability of error $P_E(x)$ is the probability of assigning $C_{+1}$ to $x \in [-2, \delta]$ plus the probability of assigning $C_{-1}$ to $x \in (\delta, 2]$. For $|x| > 2$, $P_E(x) = 0$ since $C_{-1}$ does not happen.

Therefore, we have

$$P_E(x) = \int_{-2}^{\delta} p(C_{+1}|x)\,dx + \int_{\delta}^{2} p(C_{-1}|x)\,dx = p(C_{+1})\int_{-2}^{\delta} p(x|C_{+1})\,dx + p(C_{-1})\int_{\delta}^{2} p(x|C_{-1})\,dx$$

$$= 0.6\int_{-2}^{\delta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-2)^2}{2}\right)dx + 0.4\int_{\delta}^{2} \frac{1}{4}\,dx, \quad |x| \leq 2.$$

Plugging $\delta = 2 - \sqrt{\log \frac{36}{2\pi}}$ in the above expression, we get $0.1880\ldots$ and so

$$P_{E,B}(x) = \begin{cases} 0, & |x| > 2 \\ 0.1880\ldots, & |x| \leq 2 \end{cases}$$

A different method will have a different philosophy, and therefore a different $\delta$, changing the value of $\delta$.

c) The maximum likelihood criterion swaps the Bayes criterion "Assign to x the class that maximizes $p(C_i|x)$" for "Assign to x the class that maximizes $p(x|C_i)$".

As such, the ML architecture remains the same as Bayes' with a slightly different discriminant function:

$$g_{ML}(x) = \log[p(x|C_{+1})] - \log[p(x|C_{-1})]$$

$$= \begin{cases} -\frac{1}{2}\log 2\pi - \frac{1}{2}(x-2)^2 - \log\frac{1}{4}, & |x| \leq 2 \\ +\infty, & |x| > 2 \text{ (since } C_{-1} \text{ cannot happen)} \end{cases}$$

<u>Boundary</u>. We have

$$g_{ML}(x) = 0 \Leftrightarrow -\frac{1}{2}\log 2\pi - \frac{1}{2}(x-2)^2 - \log\frac{1}{4} = 0 \quad (|x| \leq 2)$$

$$\Leftrightarrow |x-2| = \sqrt{\log\frac{16}{2\pi}} \quad \therefore x = 2 - \sqrt{\log\frac{16}{2\pi}} \approx 1.033.$$

<u>Error probability</u>. We have

$$P_{E,ML}(x) = \begin{cases} 0, & |x| > 2 \\ \int_{-2}^{1.03} 0.6 \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-2)^2}{2}\right) dx + \int_{1.03}^{2} 0.4 \times \frac{1}{4}\, dx, & |x| \leq 2 \end{cases}$$

$$= \begin{cases} 0, & |x| > 2 \\ 0.1967\ldots, & |x| \leq 2 \end{cases}$$

d) We see that the ML approach has greater probability of error. This happens because, when setting is discriminant, the priors are ignored. Should the priors be equal, ML would be equivalent to the bayesian approach.