MAE 0580/ MAC 6926
Lista 6

Notations

- $\mathcal{X} = \mathbb{R}^d$ denotes the space of *entries*, or *stimuli*;
- $\mathcal{Y} = \{-1, 1\}$ denotes the classifications labels;
- Given a vector $w = (w_0, \dots, w_d) \in \mathbb{R}^{d+1}$, we define the transformation $s_w : \mathcal{X} \to \mathbb{R}$ in the following way $s_w(x) = w_0 + \sum_{i=1}^d w_i x(i)$
- We state the **Jensen inequality** in the specific form it will be used in exercise 5.

  Proposition For any real number $p \in [0, 1]$, $r_1 \geq 0$ and $r_2 \geq 0$, the following inequality is verified:

$$\log\big(pr_1 + (1-p)r_2\big) \geq p\log(r_1) + (1-p)\log(r_2)\,.$$

Solutions

1. The statement we want to prove is obviously true for $\mathbb{P}(Y = 1|X = x) = \frac{1}{1+e^{-s_{\tilde{w}}(x)}}$, therefore it remains only to prove it for $Y = -1$. We have:

$$\mathbb{P}(Y = -1|X = x) = 1 - \mathbb{P}(Y = 1|X = x) = 1 - \frac{1}{1+e^{-s_{\tilde{w}}(x)}} = \frac{e^{-s_{\tilde{w}}(x)}}{1+e^{-s_{\tilde{w}}(x)}}$$

   Then, multiplying by $e^{s_{\tilde{w}}(x)}$ both the numerator and the denominator, we get the desired result, that is:

$$\mathbb{P}(Y = -1|X = x) = \frac{1}{1+e^{s_{\tilde{w}}(x)}}$$

2. Using the independence of the $(X_1, Y_1), \dots, (X_n, Y_n)$, we have that:

$$\begin{aligned}\frac{1}{n}\log\mathbb{P}(Y_1 = y_1, \dots Y_n = y_n|X_1 = x_1, \dots, X_n = x_n) &= \frac{1}{n}\log\prod_{m=1}^{n}\mathbb{P}(Y_m = y_m|X_m = x_m)\\ &= \frac{1}{n}\sum_{m=1}^{n}\log\mathbb{P}(Y_m = y_m|X_m = x_m)\end{aligned}$$

   And replacing the $\mathbb{P}(Y_i = y_i|X_i = x_i)$ by the formula from the previous exercise indeed gives us:

$$\frac{1}{n}\log\mathbb{P}(Y_1 = y_1, \dots Y_n = y_n|X_1 = x_1, \dots, X_n = x_n) = -\frac{1}{n}\sum_{m=1}^{n}\log\Big(1 + e^{-y_m s_{\tilde{w}}(x_m)}\Big)$$

3. Here because $w_0 = 0$ in each case, and because we are in dimension $d = 1$, the risk $\mathcal{E}_n(w)$ will look like this :

$$\mathcal{E}_n(w) = \frac{1}{n}\sum_{m=1}^{n}\log\left(1 + e^{-Y_m w_1 X_m}\right)$$

   The values obtained are the following :

- $w_1 = 1$, then $\mathcal{E}_{10}(w) \approx 0.7480745$
- $w_1 = 2$, then $\mathcal{E}_{10}(w) \approx 0.8832807$
- $w_1 = -1$, then $\mathcal{E}_{10}(w) \approx 0.7310745$

Tips

It's way faster to write few lines of code in your favorite language than to actually compute it by hand. Here a short R code to dot it:

```
x = c(-0.93, -0.90,0.64,-0.36,-0.93,-0.23, 0.59, -0.02, 0.29, 0.50)
y = c(1,1,-1,-1,-1,-1,1,1,-1,1)

w = 1
n = length(x)
Epsilon = 0

for(i in seq(1,n))
{
  Epsilon = Epsilon + log(1 + exp(-y[i] * w * x[i]))
}

Epsilon = -Epsilon/n
```

4. Using the law of large number we get that:

$$\begin{aligned}\lim_{n\to+\infty} \mathcal{E}_n(w) &= \mathbb{E}\left(\log(1+e^{-Ys_w(X)})\right) \\ &= -\sum_{y\in\{-1,1\}} \mathbb{E}\left(\log\left(\frac{1}{1+e^{-ys_w(X)}}\right)1_{Y=y}\right) \\ &= \sum_{y\in\{-1,1\}} \int_{\mathcal{X}} \mathbb{P}_X(dx)\log\left(\frac{1}{1+e^{-ys_w(x)}}\right)^{-1} \mathbb{P}\left(Y=y|X=x\right)\end{aligned}$$

Inversing the sum and the integral and using the notations of the exercise, we get the desired result.

5. To lighten the computation we will use $p(\tilde{w})$ as a short hand for $p_{\tilde{w}}(1|x)$ and $p(w)$ for $p_w(1|x)$. Translated in these terms, the inequality we want to show becomes :

$$p(\tilde{w})\log\big(p(w)\big) + (1-p(\tilde{w}))\log\big(1-p(w)\big) - \Big(p(\tilde{w})\log\big(p(\tilde{w})\big) + (1-p(\tilde{w}))\log\big(1-p(\tilde{w})\big)\Big) \leq 0$$

And the left part of the inequality can be rewritten :

$$p(\tilde{w})\log\left(\frac{p(w)}{p(\tilde{w})}\right) + \big(1-p(\tilde{w})\big)\log\left(\frac{1-p(w)}{1-p(\tilde{w})}\right)$$

And using Jensen inequality, this last expression is less or equal than :

$$\log\left(p(\tilde{w})\frac{p(w)}{p(\tilde{w})} + \big(1-p(\tilde{w})\big)\frac{1-p(w)}{1-p(\tilde{w})}\right)$$

Which reduces to $\log(1) = 0$.

6. We can see in exercise 3 that $\mathcal{E}_{10}(\tilde{w}) \approx 0.7480745$ is actually greater than $\mathcal{E}_{10}(w)$ for the w given by $w_0 = 0$ and $w_1 = -1$ (then, as stated in the exercise $\mathcal{E}_{10}(w) \approx 0.7310745$). This is obviously only true because our sample is too small ($n = 10$), and is never true asymptotically as we showed in previous exercise.