MAE 0580/ MAC 6926
Lista 6

Notação e desigualdade de Jensen

- $\mathcal{X} = \mathbb{R}^d$ designa o espaço das *entradas* ou *estímulos*;
- $\mathcal{Y} = \{-1, 1\}$ designa o conjunto das classificações;
- dado um vetor $w = (w_0, \dots, w_d) \in \mathbb{R}^{d+1}$, definimos a transformação $s_w : \mathcal{X} \to \mathbb{R}$ da seguinte maneira $s_w(x) = w_0 + \sum_{i=1}^d w_i x(i)$
- Vamos enunciar a **Desigualdade de Jensen** na forma particular que utilizaremos no exercício 5.

  Proposição. Para quaisquer números reais $p \in [0, 1]$, $r_1 \geq 0$ e $r_2 \geq 0$, vale a desigualdade

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

Exercícios.

1. Sejam X uma variável aleatória assumindo valores em $\mathcal{X} = \mathbb{R}^d$ e Y uma variável aleatória assumindo valores em $\mathcal{Y} = \{-1, 1\}$, definidas no mesmo espaço de probabilidades e tais que

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

   onde $\tilde{w} = (\tilde{w}_0, \dots, \tilde{w}_d) \in \mathbb{R}^{d+1}$ é um vetor fixado. Mostre que para todo valor $x \in \mathcal{X}$ e $y \in \{-1, +1\}$, a seguinte igualdade é verdadeira

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

2. Continuando o exercício 1, sejam $(X_1, Y_1), \dots, (X_n, Y_n)$ variáveis aleatórias i.i.d. com a mesma distribuição que (X, Y). Mostre que

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

3. Nas mesmas condições do exercício 2, dada uma amostra $(X_1, Y_1) \dots (X_n, Y_n)$, para cada sequência $w = (w_0, \dots, w_d) \in \mathbb{R}^{d+1}$, definimos o *risco* $\mathcal{E}_n(w)$ associado a um vetor $w \in \mathbb{R}^{d+1}$ qualquer da seguinte maneira

$$\mathcal{E}_n(w) = \frac{1}{n}\sum_{m=1}^n \log\Big(1 + e^{-Y_m s_w(X_m)}\Big)\,.$$

No caso em que X assume valores em $\mathbb{R}$ (isto é, $d = 1$), dada a seguinte amostra de comprimento $n = 10$

$$\begin{aligned} X_1 &= -0,93 \quad & Y_1 &= 1 \\ X_2 &= -0,90 \quad & Y_2 &= 1 \\ X_3 &= 0,64 \quad & Y_3 &= -1 \\ X_4 &= -0,36 \quad & Y_4 &= -1 \\ X_5 &= -0,93 \quad & Y_5 &= -1 \\ X_6 &= -0,23 \quad & Y_6 &= -1 \\ X_7 &= 0,59 \quad & Y_7 &= 1 \\ X_8 &= -0,02 \quad & Y_8 &= 1 \\ X_9 &= 0,29 \quad & Y_9 &= -1 \\ X_{10} &= 0,50 \quad & Y_{10} &= 1\,, \end{aligned}$$

calcule $\mathcal{E}_{10}(w)$, para os seguintes valores de $w = (w_0, w_1)$:

- $w_0 = 0$, $w_1 = 1$
- $w_0 = 0$, $w_1 = 2$
- $w_0 = 0$, $w_1 = -1$

4. Nas mesmas condições dos exercícios 1 e 2, observando que $\big(\log(1 + e^{-y_m s_w(x_m)})\big)_{m\geq 1}$ são variáveis aleatórias i.i.d. e usando a Lei dos Grandes Números, conclua que para quase toda amostra $(X_1, Y_1), (X_2, Y_2), \ldots$ vale a igualdade

$$\lim_{n\to+\infty} \mathcal{E}_n(w) = \int_{\mathcal{X}} \mathbb{P}_X(dx) \sum_{y\in\{-1,+1\}} p_{\tilde{w}}(y\,|\,x) \log\left[\frac{1}{p_w(y\,|\,x)}\right] ,$$

onde

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

$$p_w(y\,|\,x) = \frac{1}{1 + e^{-y s_w(x)}}$$

e P_X designa a lei da variável X.

5. Usando a desigualdade de Jensen, mostre que para todo $x \in \mathcal{X}$, vale a desigualdade

$$\sum_{y\in\{-1,+1\}} p_{\tilde{w}}(y\,|\,x) \log\left(\frac{1}{p_w(y\,|\,x)}\right) \geq \sum_{y\in\{-1,+1\}} p_{\tilde{w}}(y\,|\,x) \log\left(\frac{1}{p_{\tilde{w}}(y\,|\,x)}\right) .$$

Conclua que para valores de n suficientemente grandes o risco $\mathcal{E}_n(w)$ assume seu valor mínimo quando $w = \tilde{w}$.

6. Vamos verificar num caso simples a desigualdade enunciada no exercício anterior. No exercício 3, a classificação Y foi feita a partir do valor de X, usando o vetor ($\tilde{w}_0 = 0$, $\tilde{w}_1 = 1$). A desigualdade de Jensen garante que $\mathcal{E}_n(w) \geq \mathcal{E}_n(\tilde{w})$, quando w é um outro vetor qualquer, diferente de $\tilde{w}$. Verifique esse fato para alguns valores de $w \neq \tilde{w}$, usando os valores da amostra que aparece no exercício 3. Observação: para simplificar a resolução do exercício 3, o tamanho da amostra ($n = 10$) foi pequeno. Aqui teria sido melhor ter uma amostra de tamanho maior.