Exam ML 2014/2015

1. a)

$$f(x_1, x_2) = 6.5 + 0.5x_1 + 0.9x_2 + 0.1(x_1)(x_2)$$

$A \curvearrowright (1.2, 75)$; $B \curvearrowright (10.4, -15.0)$

$$f(A) = 6.5 + 0.5(1.2) + 0.9(75) + 0.1(1.2 \times 75) = 92$$

$$f(B) = 6.5 + 0.5(10.4) + 0.9(-15.0) + 0.1(10.4 \times (-15.0)) = -16,36$$

Como $f(A) = 92 > 6.5$ e $f(B) = -16.36 < 6.5$, então A e B pertencem a classes diferentes.

b) Analisando o dataset podemos verificar que existem, em grosso modo, dois clusters: cluster 1 - círculos; cluster 2 - Estrelas. Verifica-se também que há uma estrela no cluster 2, o que me leva a crer que se trata de um outlier por exemplo (mal etiquetado, p. ex.). Ora, se se utilizar o 1-NN, devido à existência daquele outlier, a previsão de novos pontos poderá ser influenciada negativamente. Assim, a minha sugestão seria utilizar um 3-NN, uma vez que, depois de avaliar os 3 "nearest neighbors", ele iria atribuir ao novo ponto a "label" mais votada. Consequentemente, a existência daquele outlier, por exemplo, não iria afetar aquela que se acredita ser a atribuição do cluster correto.

2.

a). Classifier Based on Bayes Decision Theory:

- Given x classify it to C_i if: $P(C_i|x) > P(C_j|x)$, $\forall j \neq i$

The loss matrix:

$$\begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}$$

Cost of deciding for C_1: $l_{11} P(C_1|x) + l_{21} P(C_2|x)$

cost of deciding for C_2: $l_{12} P(C_1|x) + l_{22} P(C_2|x)$

Decidir por C_1 se:

Custo de decidir por C_1 < custo de decidir por C_2

Decidir por C_2 se:

custo de decidir por C_1 > custo de decidir por C_2

Else:

Rejeitar.

- Neste caso, vamos assumir que $C_1 \to 0$ e $C_2 \to 1$.

Temos em mãos que $P(y=1|x) = 0.2$; então, então $P(C_2|x) = 0.2$

$\hookrightarrow$ então $P(y=0|x) = 0{,}8$; então $P(C_1|x) = 0.8$

custo de decidir por C_1:

Loss Matrix:

$$\begin{bmatrix} 0 & 10 \\ 10 & 0 \end{bmatrix}$$

$l_{11} P(C_1|x) + l_{21} P(C_2|x) =$

$= 0\,(0.8) + 10 \times (0{,}2) = \underline{\underline{2}}$

Com $2 < 3 < 8$

$\hookrightarrow$ A melhor decisão é decidir $\boxed{C_1}$

Rejeitar: $\underline{\underline{3}}$

custo de decidir por C_2: $10 \times 0.8 + 0 \times (0.2) = \underline{8}$

b) $P(y=1|x) = 0.4$; logo $P(C_2|x) = 0.4$

$P(y=0|x) = 0.6$; logo $P(C_1|x) = 0.6$

Custo de decidir por C_1:

$\ell_{11} P(C_1|x) + \ell_{21} P(C_2|x) = 0 \times 0.6 + 10 \times 0.4 = 4$

Custo de decidir por C_2:

$\ell_{12} P(C_1|x) + \ell_{22} P(C_2|x) = 10 \times 0.6 + 0 \times 0.4 = 6$

Rejeitar:

3 → Então, como $3 < 4 < 6$, a decisão com menor custo é rejeitar.

c) $P_1 = P(y=1|x)$, que no nosso caso é $P(C_2|x)$.

Portanto, nós prevemos $y=0$, i.e., classe C_1, se: (rej. $P_0 = C_2$)

$\ell_{11} P(C_1|x) + \ell_{21} P(C_2|x) < \ell_{12} P(C_1|x) + \ell_{22} P(C_2|x)$

Rejeitamos se:

$\ell_{11} P(C_1|x) + \ell_{21} P(C_2|x) \geq 3 \quad \wedge \quad \ell_{12} P(C_1|x) + \ell_{22} P(C_2|x) \geq 3$

Prevemos $y=1$, i.e. classe C_2 se:

$\ell_{11} P(C_1|x) + \ell_{21} P(C_2|x) > \ell_{12} P(C_1|x) + \ell_{22} P(C_2|x)$

Nota: $P(C_2|x) = P_1$; $P(C_1|x) = 1 - P_1$; $\ell_{11} = \ell_{22} = 0$; $\ell_{12} = \ell_{21} = 10$

Resolvendo:

Para prever $y=0$:

$$\cancel{0 \times (1-p_1)} + 10 p_1 < 10 \times (1-p_1) + \cancel{0 \times p_1}$$

$$10 p_1 < 10 - 10 p_1 \Leftrightarrow 20 p_1 < 10 \Leftrightarrow p_1 < 0.5$$

Para prever $y=1$

$$\cancel{0 \times (1-p_1)} + 10 \times (p_1) > 10(1-p_1) + \cancel{0 \times (p_1)}$$

$$10 p_1 > 10 - 10 p_1 \Leftrightarrow 20 p_1 > 10 \Leftrightarrow p_1 > 0.5$$

No entanto, temos o critério de rejeição:

$$\cancel{0 \times (1-p_1)} + 10 p_1 \geq 3 \quad \wedge \quad 10(1-p_1) + \cancel{0 \times (p_1)} \geq 3$$

$$10 p_1 \geq 3 \quad \wedge \quad 10 - 10 p_1 \geq 3$$

$$p_1 \geq \frac{3}{10} \quad \wedge \quad -10 p_1 \geq -7$$

$$p_1 \geq 0.3 \quad \wedge \quad 10 p_1 \leq +7$$

$$p_1 \geq 0.3 \quad \wedge \quad p_1 \leq 0.7$$

Há rejeição se: $0.3 \leq p_1 \leq 0.7$; prevê 0: $p_1 < 0.3$

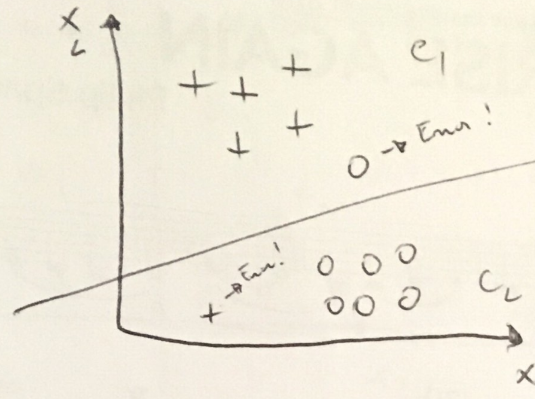
prevê 1: $p_1 > 0.7$

Logo $\theta_0 = 0.3$

$\theta_1 = 0.7$

3.

a)



(Ver: midterm 2010f-sol.pdf)

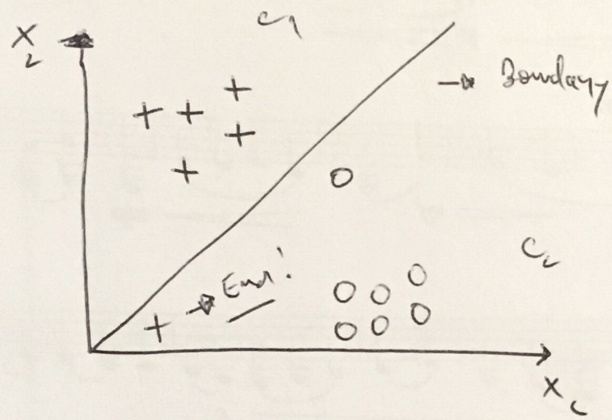
Possible boundary.

Makes 2 classification errors.

It is not a unique boundary:

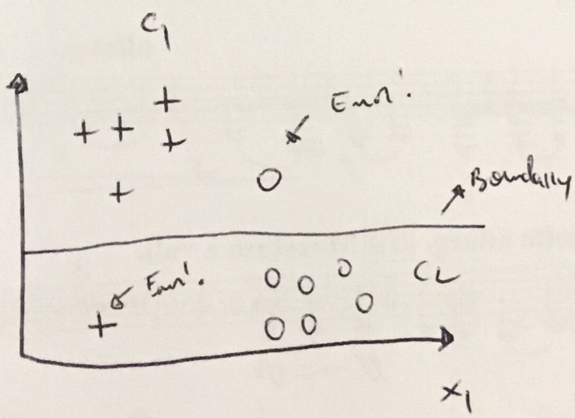
Depends on w_0, w_1 and w_2!

b)



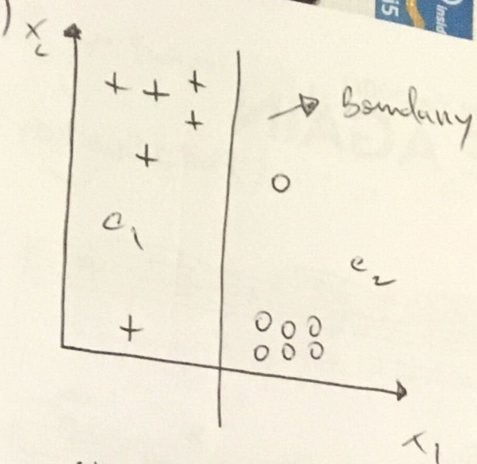
Ao regularizar w_0, bias, a "boundary" will eventually go through the origin (bias term set to zero). Pela figura, não é possível ter zero erros. No entanto, conseguimos no máximo de um erro!

c)



Ao regularizar w_1, a "boundary" deixa de depender cada vez mais de x_1 e, como tal, passa a ser cada vez mais horizontal; o training set poderá ser inclusive apenas com dois erros de classificação. Logo não se altera em relação ao inicial, mas piora em relação ao anterior.

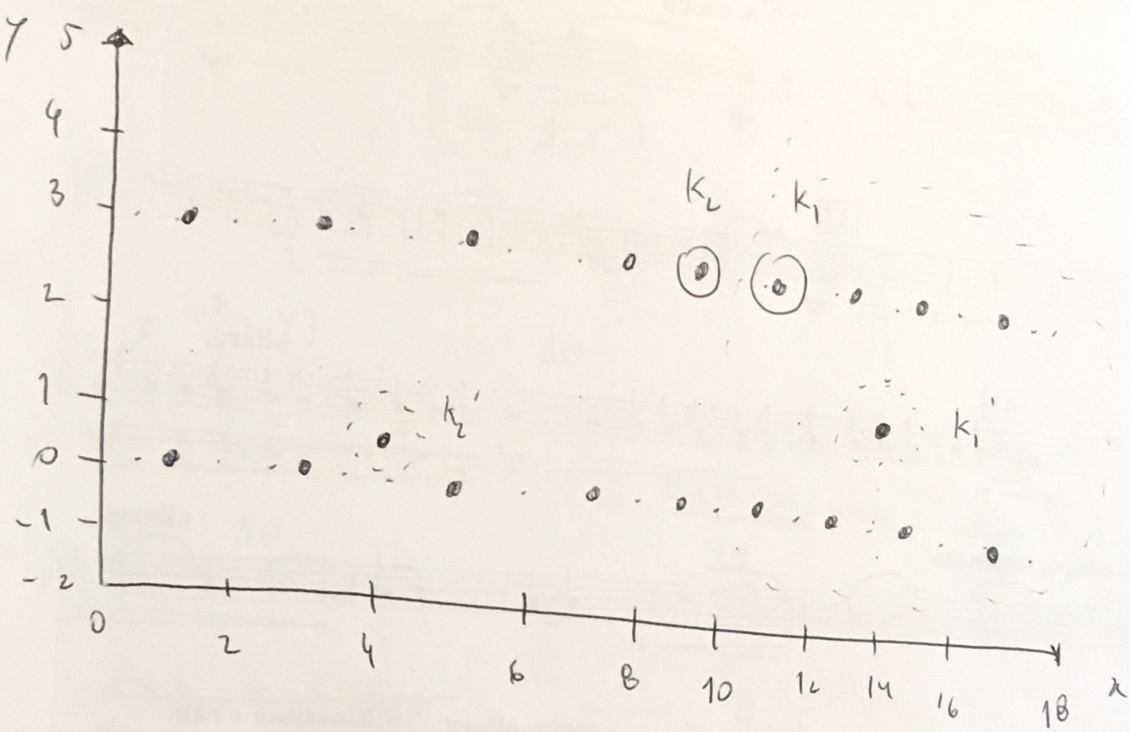
d)



No error:

Ao regularizar w_2, a "boundary" resultante irá depender cada vez mais do valor de x_1 e, como tal, irá tornar-se cada vez mais vertical. Para um λ muito grande, o erro de training diminui ao ponto de ser possível separar as classes com erro = 0.

4.



Com esta inicialização de cluster, todos os pontos com abcissa $x \le 8$ irão pertencer ao cluster K_2. Por sua vez, todos os pontos com abcissa $x > 8$ irão pertencer ao cluster K_1. Para calcular os novos "centroids", há que efetuar o cálculo das médias de cada coordenada pertencente aos pontos de cada cluster:

Novo centróide K_1:

$K_1 \rightarrow (x, y)$; $n = 8$

$$x = \frac{11\times2 + 13\times2 + 15\times2 + 17\times2}{8} = \frac{22+26+30+34}{8} = \frac{112}{8} = 14$$

$$y = \frac{4\times3 + 4\times0}{8} = \frac{12}{8} = \frac{3}{2} = 1,5$$

Logo,

$K_1 \rightarrow (14 ; 1,5)$

$K_2 \rightarrow (x, y)$; $n = 10$

$$x = \frac{9\times2 + 7\times2 + 5\times2 + 3\times2 + 1\times2}{10} = \frac{18+14+10+6+2}{10} = \frac{42}{10} = \frac{4,2}{1} = 4,2$$

$$y = \frac{5\times3 + 5\times0}{10} = \frac{5}{10} = 0,5$$

Logo, $K_2 \rightarrow (4,2 ; 0.5)$

tão simples!

5. $\boxed{P(x_2 \mid x_1, x_4, x_3)}$. $P(x_2 \mid x_1) \times P(x_2 \mid x_4) \times P(x_2, x_3)$

$P(x_1) = [0.5 \quad 0 \quad 0 \quad 0.5]$

$P(x_2) = ?$

$P(x_3) = [1 \quad 0 \quad 0]$

$P(x_4) = [0 \quad 1]$

$P(A \cap B) = P(A \mid B) \cdot P(B)$

$P(A_4 \cap A_3 \cap A_2 \cap A_1) = P(A_4 \mid A_3 \cap A_2 \cap A_1) \cdot P(A_3 \mid A_2 \cap A_1)$

$\cdot P(A_2 \cap A_1) \cdot P(A_1)$

$P(x_1) \cdot P(x_2 \mid x_1) = \underset{1 \times 4}{[0.5 \quad 0 \quad 0 \quad 0.5]} \underset{4 \times 2}{\begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \\ 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}}$

$= [0.65 \quad 0.15]$

$p(x_3) \cdot p$

Sabendo a base fiz as seguintes probabilidades

salmon : 0.65

sea bass : 0.15

Sabendo que é thin :

salmon : 0.6

sea bass : 0.05

Sabendo que é light :

salmon : 0.33

sea bass : 0.8

Então :

salmon : 0.65 x 0.6 x 0.33 0,1287

sea bass : 0.15 x 0.005 x 0.8 : 0,006

↳ classifica como salmon!

check

Bayesian Networks

6.

Bays formula: $P(c_j|x) = \frac{P(x|c_j)\,P(c_j)}{P(x)}$

$$P(x) = \sum_{j=1}^{2} (P x|c_j)\,P(c_j)$$

$L(\theta) = P(x_1, \ldots, x_n; \theta)$

$$\theta^* = \arg\max_\theta L(\theta) = \arg\max_\theta \log L(\theta)$$

a)

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} \theta_2 - \theta_1\theta_2 & 1+\theta_1\theta_2 - \theta_1 - \theta_2 \\ \theta_1 - \theta_1\theta_2 & \theta_1\theta_2 \end{bmatrix}$$

$V_{11} = P(y=0 | x=0, \theta_2)\, P(x=0|\theta_1) =$

$= \theta_2 (1-\theta_1) = \theta_2 - \theta_1\theta_2$

$V_{12} = P(y=1 | x=0, \theta_2)\, P(x=0|\theta_1) = (1-\theta_2)(1-\theta_1) =$

$= 1 - \theta_1 - \theta_2 + \theta_1\theta_2$

$= 1 + \theta_1\theta_2 - \theta_1 - \theta_2$

$v_{21} = p(y=0 \mid x=1, \theta_2)\, p(x=1 \mid \theta_1) = (1-\theta_2)(\theta_1) = \theta_1 - \theta_1\theta_2$

$v_{22} = p(y=1 \mid x=1, \theta_2)\, p(x=1 \mid \theta_1) = \theta_2(\theta_1) = \theta_1\theta_2$

5).

$$L(\theta) = (\theta_1\theta_2)(\theta_1 - \theta_1\theta_2)(\theta_2 - \theta_1\theta_2)(\theta_1 - \theta_1\theta_2)(\theta_1\theta_2)$$

$$(\theta_2 - \theta_1\theta_2)(1 + \theta_1\theta_2 - \theta_1 - \theta_2) =$$

$$= (\theta_1^2\theta_2 - \theta_1^2\theta_2^2)(\theta_2\theta_1 - \theta_1\theta_2^2 - \theta_1^2\theta_2 + \theta_1^2\theta_2^2)(\theta_1\theta_2^2 - \theta_1^2\theta_2^2)$$

$$(1 + \theta_1\theta_2 - \theta_1 - \theta_2) =$$

$$=$$

$\boxed{\text{Verifica 4}}$