Aprendizagem Homework 3 Part 1

Daniel Castro 87644
João Tiago Aparício 97155

I) Clustering

$x_1 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$ $\quad x_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ $\quad x_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\qquad N = \dfrac{e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}}{2\pi^{K/2}\sqrt{|\Sigma|}}$

$K = 2$

$\mu_1 = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$ $\quad \mu_2 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$ $\qquad \Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\quad \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$\pi_1 = p(c_1 = 1) = 0.7$ $\qquad \pi_2 = p(c_2 = 1) = 0.3$

1st Epoch

E-step

$\boxed{x_1 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}}$

C = 1 (Cluster 1)

- Prior: $p(C=1) = 0.7$
- Likelihood: $p(x_1 \mid C=1) = \underbrace{\dfrac{e^{-\frac{1}{2}(x_1-\mu_1)^T \Sigma_1^{-1}(x_1-\mu_1)}}{(2\pi)^{K/2}\sqrt{|\Sigma_1|}}}_{N(x_1 \mid \mu_1, \Sigma_1) =} = \dfrac{e^{-\frac{1}{2}\left(\begin{pmatrix} 2 \\ 4 \end{pmatrix}-\begin{pmatrix} 0 \\ 4 \end{pmatrix}\right)^T \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1}\left(\begin{pmatrix} 2 \\ 4 \end{pmatrix}-\begin{pmatrix} 0 \\ 4 \end{pmatrix}\right)}}{2\pi^{2/2}\sqrt{\left|\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right|}} =$

$= \dfrac{e^{\begin{pmatrix} -1 \\ 0 \end{pmatrix}^T \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix}}}{2\pi} = \dfrac{e^{-2}}{2\pi} \cong 0,0215$

- Joint Probability: $p(C=1, x_1) = p(C=1) . p(x_1 \mid C=1) = 0.7 \times 0.0215 \cong 0.0151$

C = 2 (Cluster 2)

- Prior: $p(C=2) = 0.3$
- Likelihood: $p(x_1 \mid C=2) = N(x_1 \mid \mu_2, \Sigma_2) = \dfrac{e^{-\frac{1}{2}(x_1-\mu_2)^T \Sigma_2^{-1}(x_1-\mu_2)}}{(2\pi)^{\frac{2}{2}}\sqrt{|\Sigma_2|}} \cong 7.226 \times 10^{-6}$
- Joint Probability: $P(C=2, x_1) = p(C=2) . P(x_1 \mid C=2) = 0.3 \times 7,226 \times 10^{-6} \cong 2,168 \times 10^{-6}$

Normalized Posteriors

- $C = 1$: $p(C=1 \mid x_1) = \dfrac{P(C=1, x_1)}{P(C=1, x_1) + P(C=2, x_1)} = \dfrac{0.0151}{0.0151 + 2,168 \times 10^{-6}} \simeq 0.999856$

- $C = 2$: $p(C=2 \mid x_1) = \dfrac{P(C=2, x_1)}{P(C=1, x_1) + P(C=2, x_1)} = \dfrac{2,168 \times 10^{-6}}{0.0151 + 2,168 \times 10^{-6}} \simeq 0.000144$

$$\boxed{x_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}}$$

$C = 1$

- Prior : $P(C=1) = 0.7$
- Likelihood: $P(x_2 \mid C=1) = N(x_2 \mid \mu_1, \Sigma_1) = N\left(\begin{pmatrix} 4 \\ 2 \end{pmatrix} \middle| \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \simeq 7.226 \times 10^{-6}$
- Joint Probability: $P(C=1, x_2) = P(C=1)\, P(\cancel{C=1 \mid x_2}\ x_2 \mid C=1) = 0.7 \times 7.226 \times 10^{-6} \simeq 5.0582 \times 10^{-6}$

$C=2$

- Prior : $P(C=2) = 0.3$
- Likelihood : $P(x_2 \mid C=2) = N(x_2 \mid \mu_2, \Sigma_2) = N\left(\begin{pmatrix} 4 \\ 2 \end{pmatrix} \middle| \begin{pmatrix} 4 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \simeq 0.0215$
- Joint Probability: $P(C=2, x_2) = P(C=2)\, P(x_2 \mid C=2) = 0.3 \times 0.0215 \simeq 0.00645$

Normalized Posteriors

- $C = 1$: $P(C=1 \mid x_2) = \dfrac{P(C=1, x_2)}{P(C=1, x_2) + P(C=2, x_2)} = \dfrac{5.0582 \times 10^{-6}}{5.0582 \times 10^{-6} + 0.00645} \simeq 0.000784$
- $C = 2$: $P(C=2 \mid x_2) = \dfrac{P(C=2, x_2)}{P(C=2, x_2) + P(C=2, x_2)} = \dfrac{0.00645}{5.0582 \times 10^{-6} + 0.00645} \simeq 0.999216$

$$\boxed{x_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}}$$

$C = 1$

- Prior : $P(C=1) = 0.7$
- Likelihood: $P(x_3 \mid C=1) = N(x_3 \mid \mu_1, \Sigma_1) = N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \middle| \begin{pmatrix} 0 \\ 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \simeq 5.339 \times 10^{-5}$
- Joint Probability : $P(C=1, x_3) = P(C=1)\, P(x_3 \mid C=1) = 0.7 \times 5.339 \times 10^{-5} \simeq 3.737 \times 10^{-5}$

$C = 2$

- Prior : $P(C=2) = 0.3$
- Likelihood : $P(x_3 \mid C=2) = N(x_3 \mid \mu_2, \Sigma_2) = N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \middle| \begin{pmatrix} 4 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \simeq 5.339 \times 10^{-5}$
- Joint Probability : $P(C=2, x_3) = P(C=2)\, P(x_3 \mid C=2) = 0.3 \times 5.339 \times 10^{-5} \simeq 1.602 \times 10^{-5}$

Normalized Posteriors

- $C = 1$: $P(C=1 \mid x_3) = \dfrac{P(C=1, x_3)}{P(C=1, x_3) + P(C=2, x_3)} = \dfrac{3.737 \times 10^{-5}}{3.737 \times 10^{-5} + 1.602 \times 10^{-5}} \simeq 0.7$
- $C=2$: $P(C=2 \mid x_3) = \dfrac{P(C=2, x_3)}{P(C=1, x_3) + P(C=2, x_3)} = \dfrac{1.602 \times 10^{-5}}{3.737 \times 10^{-5} + 1.602 \times 10^{-5}} \simeq 0.3$

Aprendizagem Homework 3 Part 2

M-Step

$N=3$ number of points in the data

$$\mu_c = \frac{\sum_{m=1}^{3N} p(C=c \mid x_m)\, x_m}{\sum_{m=1}^{3N} p(C=c \mid x_m)}$$ mean

$$\Sigma_{ij}^{c} = \frac{\sum_{m=1}^{3N} p(C=c \mid x_m)(x_{m_i} - \mu_{c_i})(x_{m_j} - \mu_{m_j})}{\sum_{m=1}^{3N} p(C=c \mid x_m)}$$ covariance matrix

$$p(C=c) = \frac{\sum_{m=1}^{N} p(C=c \mid x_m)}{\sum_{l=1}^{K}\sum_{m=1}^{N} p(C=l \mid x_m)}$$ = since the probabilities are complemented, ~~the~~ i.e., the sum of all clusters probabilies is 1, then $= \frac{\sum_{m=1}^{N=3} p(C=c \mid x_m)}{\underbrace{N=3}_{\text{number of points in the data}}}$

$C=1$

$$\mu_1 = \frac{p(C=1 \mid x_1)x_1 + p(C=1 \mid x_2)x_2 + p(C=1 \mid x_3)x_3}{p(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)} =$$

$$= \frac{0.999856 \times \begin{pmatrix}2\\4\end{pmatrix} + 7.84\times10^{-4}\begin{pmatrix}4\\2\end{pmatrix} + 0.7 \times \begin{pmatrix}0\\0\end{pmatrix}}{0.999856 + 7.84\times10^{-4} + 0.7} \simeq \begin{pmatrix}1{,}178\\2{,}353\end{pmatrix}$$

$$\Sigma_{1_{1,1}} = \frac{p(C=1 \mid x_1)(x_{1_1} - \mu_{1_1})(x_{1_1} - \mu_{1_1}) + p(C=1 \mid x_2)(x_{2_1} - \mu_{1_1})(x_{2_1} - \mu_{1_1}) + p(C=1 \mid x_3)(x_{3_1} - \mu_{1_1})(x_{3_1} - \mu_{1_1})}{p(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)}$$

$$= \frac{0.999856 \times (2-1{,}178)(2-1{,}178) + 7.84\times10^{-4} \times (4-1{,}178)(4-1{,}178) + 0.7(0-1{,}178)(0-1{,}178)}{0.999856 + 7.84\times10^{-4} + 0.7} = 0.972$$

$$\Sigma_{1_{1,2}} = \Sigma_{1_{2,1}} = \frac{p(C=1 \mid x_1)(x_{1_1} - \mu_{1_1})(x_{1_2} - \mu_{1_2}) + p(C=1 \mid x_2)(x_{2_1} - \mu_{1_1})(x_{2_2} - \mu_{1_2}) + p(C=1 \mid x_3)(x_{3_1} - \mu_{1_1})(x_{3_2} - \mu_{1_2})}{p(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)}$$

$$= \frac{0.99856 \times (2-1.178)(4-2{,}353) + 7.84\times10^{-4} \times (4-1{,}178)(2-2{,}353) + 0.7(0-1{,}178)(0-2{,}353)}{0.999856 + 7.84\times10^{-4} + 0.7} = 1{,}936$$

$$\Sigma_{1_{2,2}} = \frac{p(C=1 \mid x_1)(x_{1_2} - \mu_{1_2})^2 + p(C=1 \mid x_2)(x_{2_2} - \mu_{1_2})^2 + p(C=1 \mid x_3)(x_{3_2} - \mu_{1_2})^2}{p(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)} = 3.874$$

$$\Sigma_1 = \begin{pmatrix} 0.972 & 1.936 \\ 1.936 & 3.874 \end{pmatrix}$$

$\underline{C=2}$

$$\mu_2 = \frac{p(C=2|x_1)x_1 + p(C=2|x_2)x_2 + p(C=2|x_3)x_3}{p(C=2|x_1) + p(C=2|x_2) + p(C=2|x_3)} \simeq \begin{pmatrix} 3.076 \\ 1.538 \end{pmatrix}$$

$$\Sigma_{2_{1,1}} = \frac{p(C=2|x_1)(x_{1_1}-\mu_{2_1})^2 + p(C=2|x_2)(x_{2_1}-\mu_{2_1})^2 + p(C=2|x_3)(x_{3_1}-\mu_{2_1})^2}{p(C=2|x_1) + p(C=2|x_2) + p(C=2|x_3)} = 2.841$$

$$\Sigma_{2_{1,2}} = \Sigma_{2_{2,1}} = \frac{p(C=2|x_1)(x_{1_1}-\mu_{2_1})(x_{1_2}-\mu_{2_2}) + p(C=2|x_2)(x_{2_1}-\mu_{2_1})(x_{2_2}-\mu_{2_2}) + p(C=2|x_3)(x_{3_1}-\mu_{2_1})(x_{3_2}-\mu_{2_2})}{p(C=2|x_1) + p(C=2|x_2) + p(C=2|x_3)}$$

$= 1.420$

$$\Sigma_{2_{2,2}} = \frac{p(C=2|x_1)(x_{1_2}-\mu_{2_2})^2 + p(C=2|x_2)(x_{2_2}-\mu_{2_2})^2 + p(C=2|x_3)(x_{3_2}-\mu_{2_2})^2}{p(C=2|x_1) + p(C=2|x_2) + p(C=2|x_3)} = 0.711$$

$$\Sigma_2 = \begin{pmatrix} 2.841 & 1.42 \\ 1.42 & 0.711 \end{pmatrix}$$

$$p(C=1) = \frac{p(C=1|x_1) + p(C=1|x_2) + p(C=1|x_3)}{3} = 0.567$$

$$p(C=2) = \frac{p(C=2|x_1) + p(C=2|x_2) + p(C=2|x_3)}{3} = 0.433$$

$\mu_1 = \begin{pmatrix} 1.178 \\ 2.353 \end{pmatrix}$ $\quad \mu_2 = \begin{pmatrix} 3.076 \\ 1.538 \end{pmatrix}$ $\quad \Sigma_1 = \begin{pmatrix} 0.972 & 1.936 \\ 1.936 & 3.874 \end{pmatrix}$ $\quad \Sigma_2 = \begin{pmatrix} 2.841 & 1.42 \\ 1.42 & 0.711 \end{pmatrix}$

$|\Sigma_1| = 0.0174$ $\quad |\Sigma_2| = 0.00355$

$p(C=1) = 0.567$ $\quad p(C=2) = 0.433$

$\Sigma_1^{-1} = \begin{pmatrix} 222.235 & -111.06 \\ -111.06 & 55.760 \end{pmatrix}$ $\quad \Sigma_2^{-1} = \begin{pmatrix} 200.225 & -399.887 \\ -399.887 & 800.056 \end{pmatrix}$

“Quando chegares ao cimo de uma montanha continua a subir.”

(Expressão Zen)

Aprendizagem Homework 3 Part 3

2nd Epoch

E-step

$\boxed{x_1 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}}$

C=1

- Prior: $p(c=1) = 0.567$
- Likelihood: $p(x_1 | c=1) = N(x_1 | \mu_1, \Sigma_1) = N\left(\begin{pmatrix} 2 \\ 4 \end{pmatrix} \middle| \begin{pmatrix} 1.178 \\ 2.353 \end{pmatrix} \begin{pmatrix} 0.972 & 1.936 \\ 1.936 & 3.874 \end{pmatrix}\right) \simeq 0.849$
- Joint Probability: $p(c=1 | x_1) = p(c=1) p(x_1 | c=1) = 0.567 \times 0.849 = 0.481$

C=2

- Prior: $p(c=2) = 0.433$
- Likelihood: $p(x_1 | c=2) = N(x_1 | \mu_2, \Sigma_2) = \cancel{N((\ldots))} N\left(\begin{pmatrix} 2 \\ 4 \end{pmatrix} \middle| \begin{pmatrix} 3.076 \\ 1.538 \end{pmatrix} \begin{pmatrix} 2.841 & 1.42 \\ 1.42 & 0.711 \end{pmatrix}\right) \simeq 0$
- Joint Probability: $p(c=2 | x_1) = p(c=2) p(x_1 | c=2) = 0.433 \times 0 = 0$

Normalized Posteriors

- $C=1$: $p(c=1 | x_1) = \dfrac{p(c=1, x_1)}{p(c=1, x_1) + p(c=2, x_1)} = 1$
- $C=2$: $p(c=2 | x_1) = 0$

$\boxed{x_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}}$

C=1

- Prior: $p(c=1) = 0.567$
- Likelihood: $p(x_2 | c=1) = N(x_2 | \mu_1, \Sigma_1) \simeq 0$
- Joint Probability: $p(c=1 | x_2) = p(c=1) p(x_2 | c=1) = 0$

C=2

- Prior: $p(c=2) = 0.433$
- Likelihood: $p(x_2 | c=2) = N(x_2 | \mu_2, \Sigma_2) = 2.298$
- Joint Probability: $p(c=2 | x_2) = p(c=2) p(x_2 | c=2) = 0.433 \times 2.298 = 0.995$

Normalized Posteriors

$C=1$: $p(c=1 | x_2) = 0$

$C=2$: $p(c=2 | x_2) = 1$

$\boxed{x_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}}$

$\underline{C=1}$

- Prior: $P(C=1) = 0.567$
- Likelihood: $P(x_3 \mid C=1) = N(x_3 \mid \mu_1, \Sigma_1) = 0.5890$
- Joint Probability: $P(C=1, x_3) = P(C=1)\, p(x_3 \mid C=1) = 0.567 \times 0.59 = 0.335$

$\underline{C=2}$

- Prior: $P(C=2) = 0.433$
- Likelihood: $P(x_3 \mid C=2) = N(x_3 \mid \mu_2, \Sigma_2) = 0.505$
- Joint Probability: $P(C=2, x_3) = P(C=2)\, P(x_3 \mid C=2) = 0.433 \times 0.505 = 0.219$

Normalized Posteriors

- $C=1$: $$P(C=1 \mid x_3) = \frac{P(C=1, x_3)}{P(C=1, x_3) + P(C=2, x_3)} = \frac{0.335}{0.335 + 0.219} = 0.605$$
- $C=2$: $$P(C=2 \mid x_3) = \frac{P(C=2, x_3)}{P(C=2, x_3) + P(C=2, x_3)} = \frac{0.219}{0.335 + 0.219} = 0.395$$

M-step

$\underline{C=1}$

$$\mu_1 = \frac{p(C=1 \mid x_1)x_1 + p(C=1 \mid x_2)x_2 + p(C=1 \mid x_3)x_3}{p(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)} = \begin{pmatrix} 1.246 \\ 2.492 \end{pmatrix}$$

$$\Sigma_{1_{1,1}} = \frac{p(C=1 \mid x_1)(x_{1_1} - \mu_{1_1})^2 + p(C=1 \mid x_2)(x_{2_1} - \mu_{1_1})^2 + p(C=1 \mid x_3)(x_{3_1} - \mu_{1_1})^2}{p(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)} = 0.944$$

$$\Sigma_{1_{1,2}} = \Sigma_{1_{2,1}} = \frac{p(C=1 \mid x_1)(x_{1_1} - \mu_{1_1})(x_{1_2} - \mu_{1_2}) + p(C=1 \mid x_2)(x_{2_1} - \mu_{1_1})(x_{2_2} - \mu_{1_2}) + p(C=1 \mid x_3)(x_{3_1} - \mu_{1_1})(x_{3_2} - \mu_{1_2})}{p(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)}$$

$= 1.879$

$$\Sigma_{1_{2,2}} = \frac{p(C=1 \mid x_1)(x_{1_2} - \mu_{1_2})^2 + p(C=1 \mid x_2)(x_{2_2} - \mu_{1_2})^2 + p(C=1 \mid x_3)(x_{3_2} - \mu_{1_2})^2}{p(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)} = \cancel{1.879}\ 3.758$$

$$\Sigma_1 = \begin{pmatrix} 0.944 & 1.879 \\ 1.879 & 3.758 \end{pmatrix}$$

$$P(C=1) = \frac{P(C=1 \mid x_1) + p(C=1 \mid x_2) + p(C=1 \mid x_3)}{3} = 0.5835$$

Aprendizagem Homework 3 Part 4

$C = 2$

$$\mu_2 = \frac{p(C=2|x_1)x_1 + p(C=2|x_2)x_2 + p(C=2|x_3)x_3}{p(C=2|x_1) + p(C=2|x_2) + p(C=2|x_3)} = \begin{pmatrix} 2.867 \\ 1.434 \end{pmatrix}$$

$$\Sigma_{2_{1,1}} = \frac{p(C=2|x_1)(x_{1_1} - \mu_{2_1})^2 + p(C=2|x_2)(x_{2_1} - \mu_{2_1})^2 + p(C=2|x_3)(x_{3_1} - \mu_{2_1})^2}{p(C=2|x_1) + p(C=2|x_2) + p(C=2|x_3)} = 3.248$$

$$\Sigma_{2_{1,2}} = \Sigma_{2_{2,1}} = \frac{p(C=2|x_1)(x_{1_1} - \mu_{2_1})(x_{1_2} - \mu_{2_2}) + p(C=2|x_2)(x_{2_1} - \mu_{2_1})(x_{2_2} - \mu_{2_2}) + p(C=2|x_3)(x_{3_1} - \mu_{2_1})(x_{3_2} - \mu_{2_2})}{p(C=2|x_1) + p(C=2|x_2) + p(C=2|x_3)} = 1.624$$

$$\Sigma_{2_{2,2}} = 0.812$$

$$\Sigma_2 = \begin{pmatrix} 3.248 & 1.624 \\ 1.624 & 0.812 \end{pmatrix}$$

$$P(C=2) = 0.465$$

Aprendizagem Homework 3 Part 5

II) RBF

$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $x_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ $x_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ $x_4 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

$t_1 = 1$ $t_2 = 0$ $t_3 = 0$ $t_4 = 1$

$c_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $c_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$

$K = 2$ $\sigma = 1$

K-means

1st Epoch

(using Euclidean distance)

$d_{x,c_1} = [0, 1, 1, \sqrt{2}]$

$d_{x,c_2} = [1, \sqrt{2}, 0, 1]$

$c_1 = \dfrac{\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}}{2} = \begin{pmatrix} 0 \\ -0.5 \end{pmatrix}$

$c_2 = \dfrac{\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \end{pmatrix}}{2} = \begin{pmatrix} -1 \\ -0.5 \end{pmatrix}$

Does not converge because

$\begin{pmatrix} 0 \\ -0.5 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\begin{pmatrix} -1 \\ -0.5 \end{pmatrix} \neq \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ ›

2nd Epoch

$d_{x,c_1} = [0.5, 0.5, \sqrt{1.25}, \sqrt{1.25}]$

$d_{x,c_2} = [\sqrt{1.25}, \sqrt{1.25}, 0.5, 0.5]$

$c_1 = \dfrac{\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}}{2} = \begin{pmatrix} 0 \\ -0.5 \end{pmatrix}$

$c_2 = \dfrac{\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \end{pmatrix}}{2} = \begin{pmatrix} -1 \\ -0.5 \end{pmatrix}$

Converges and al
it's two epochs d

$\begin{pmatrix} 0 \\ -0.5 \end{pmatrix} = \begin{pmatrix} 0 \\ -0.5 \end{pmatrix}$ ✓

$\begin{pmatrix} -1 \\ -0.5 \end{pmatrix} = \begin{pmatrix} -1 \\ -0.5 \end{pmatrix}$ ✓

$\eta = 1$ $W = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $b = [1]$ $\sigma = 1$

$\phi_k = \phi(\|x - c_k\|) = e^{-\frac{\|x - c_k\|^2}{2 \times \sigma^2}}$

$\text{Sigma}(x) = \dfrac{e^x}{1+e^x}$

$F(x) = \sum_{k=1}^{K} \left(w_k \cdot \phi(\|x - c_k\|)\right) + b$

Update weight and bias formulas

$w = w + \eta \cdot \sum_{n=1}^{N} (t_n - o_n) \cdot \phi_n$

$b = b + \eta \cdot$

NOTA: Devido a um script criado por nós e que está em anexo, os cálculos seguintes estão simplificados, exceto na 1ª iteração.

$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ **1st Epoch**

$$\phi_{K=1} = \phi(\|x_1 - c_1\|) = e^{-\frac{\|x_1 - c_1\|^2}{2 \cdot 1^2}} = e^{-0.125}$$

$$\phi_{K=2} = \phi(\|x_1 - c_2\|) = e^{-\frac{\|x_1 - c_2\|^2}{2 \cdot \sigma^2}} = e^{-\frac{\left\|\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ -0.5 \end{pmatrix}\right\|^2}{2 \cdot 1^2}} = e^{-0.625}$$

$$F(x_1) = \phi(\|x_1 - c_1\|) \cdot W_1 + \phi(\|x_1 - c_2\|) \cdot W_2 + b =$$

$$= e^{-0.125} \cdot 1 + e^{-0.625} \cdot 1 + 1 =$$

$$\text{Sigma}(F(x_1)) = \frac{e^{F(x)}}{1 + e^{F(x)}} = \frac{e^{e^{-0.125} + e^{-0.625} + 1}}{1 + e^{e^{-0.125} + e^{-0.625} + 1}} \approx 0.918$$

$$W = W + \eta \times (t - \text{sigma}(F(x_1))) \times \begin{bmatrix} \phi_{K=1} \\ \phi_{K=2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \times (1 - 0.918) \times \begin{bmatrix} e^{-0.125} \\ e^{-0.625} \end{bmatrix} = \begin{bmatrix} 1.072 \\ 1.044 \end{bmatrix}$$

$$b = b + ($$

$x_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

$\phi_1 = 0.882$ $\phi_2 = 0.535$

$F(x_2) = 2.587$

$\text{Sigma}(F(x_2)) = 0.930$

$W = \begin{bmatrix} 0.251 \\ 0.546 \end{bmatrix}$ $b = 0.152$

$x_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$

$\phi_1 = 0.535$ $\phi_2 = 0.882$

$F(x_3) = 0.768$

$\text{Sigma}(F(x_3)) = 0.683$

$W = \begin{bmatrix} -0.114 \\ -0.057 \end{bmatrix}$ $b = -0.531$

$x_4 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

$\phi_1 = 0.535$ $\phi_2 = 0.882$

$F(x_4) = -0.643$

$\text{Sigma}(F(x_4)) = 0.345$

$W = \begin{bmatrix} 0.236 \\ 0.521 \end{bmatrix}$

$b = 0.124$

Aprendizagem Homework 3 Part 6

2^{nd} Epoch

$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\phi_1 = 0.882 \qquad \phi_2 = 0.535$

$F(x_1) = 0.612 \qquad \text{sigma}(F(x_1)) = 0.648$

$W = \begin{bmatrix} 0.547 \\ 0.710 \end{bmatrix} \qquad b = 0.476$

$x_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

$\phi_1 = 0.882 \qquad \phi_2 = 0.535$

$F(x_2) = 1.338 \qquad \text{sigma}(F(\cancel{x_3}x_2)) = 0.792$

$W = \begin{bmatrix} -0.152 \\ 0.286 \end{bmatrix} \qquad b = -0.317$

$x_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$

$\phi_1 = 0.535 \qquad \phi_2 = 0.882$

$F(x_3) = -0.146 \qquad \text{sigma}(F(x_3) = 0.464$

$W = \begin{bmatrix} -0.400 \\ -0.123 \end{bmatrix} \qquad b = -0.780$

$x_4 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

$\phi_1 = 0.535 \qquad \phi_2 = 0.882$

$F(x_4) = -1.103 \qquad \text{sigma}(F(x_4)) = 0.24$

$W = \begin{bmatrix} 0.00159 \\ 0.539 \end{bmatrix} \qquad b = -0.0292$

3^{rd} Epoch

$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\phi_1 = 0.882$ $\phi_2 = 0.535$

$F(x_1) = 0.261$ $\text{sigma}(F(x_1)) = 0.565$

$W = \begin{bmatrix} 0.386 \\ 0.772 \end{bmatrix}$ $b = 0.406$

$x_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

$\phi_1 = 0.882$ $\phi_2 = 0.535$

$F(x_2) = 1.159$ $\text{sigma}(F(x_2)) = 0.761$

$W = \begin{bmatrix} -0.286 \\ 0.365 \end{bmatrix}$ $b = -0.355$

$x_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$

$\phi_1 = 0.535$ $\phi_2 = 0.882$

$F(x_3) = -0.187$ $\text{sigma}(F(x_3)) = 0.453$

$W = \begin{bmatrix} -0.529 \\ -0.0355 \end{bmatrix}$ $b = -0.809$

$x_4 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

$\phi_1 = 0.535$ $\phi_2 = 0.882$

$F(x_4) = -1.123$

$W = \begin{bmatrix} -0.125 \\ 0.631 \end{bmatrix}$ $b = -0.0542$

b)

Para o ponto $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, o output da RBF network é 0.48:

$\phi_1 = 0.197$ $\phi_2 = 0.0440$

$F\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = -0.0511$ $\text{sigma}\left(F\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)\right) = 0.487$

Visto que o target só pode ser 0 ou 1, 0.487 está mais pert
Logo o ponto $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ tem $t = 0$.

Machine Learning Homework

Daniel Castro 87644
João Lago Aparício 97155

III PCA

$X = \left(x_1 = \begin{pmatrix}2\\4\end{pmatrix}, x_2 = \begin{pmatrix}4\\2\end{pmatrix}, x_3 = \begin{pmatrix}0\\4\end{pmatrix}, x_4 = \begin{pmatrix}1\\5\end{pmatrix}\right)$

a)

$$C_{ij} = \frac{\sum_{k=1}^{m} \left(x_i^{(k)} - m_i\right)\left(x_j^{(k)} - m_j\right)}{m-1} \qquad C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$m_1 = \frac{(2+4+0+1)}{4} = \frac{7}{4}$$

$$m_2 = \frac{(4+2+4+5)}{4} = \frac{15}{4}$$

$$C_{11} = \frac{\left(2-\frac{7}{4}\right)^2 + \left(4-\frac{7}{4}\right)^2 + \left(0-\frac{7}{4}\right)^2 + \left(1-\frac{7}{4}\right)^2}{3} = \frac{35}{12}$$

$$C_{12} = C_{21} = \frac{\left(2-\frac{7}{4}\right)\cdot\left(4-\frac{15}{4}\right) + \left(4-\frac{7}{4}\right)\cdot\left(2-\frac{15}{4}\right) + \left(0-\frac{7}{4}\right)\left(4-\frac{15}{4}\right) + \left(1-\frac{7}{4}\right)\left(5-\frac{15}{4}\right)}{3} = -\frac{7}{4}$$

$$C_{22} = \frac{\left(4-\frac{15}{4}\right)^2 + \left(2-\frac{15}{4}\right)^2 + \left(4-\frac{15}{4}\right)^2 + \left(5-\frac{15}{4}\right)^2}{3} = \frac{19}{12}$$

$$C = \begin{bmatrix} \frac{35}{12} & -\frac{7}{4} \\ -\frac{7}{4} & \frac{19}{12} \end{bmatrix} = \begin{bmatrix} 2.91667 & -1{,}75 \\ -1{,}75 & 1{,}58333 \end{bmatrix}$$

$$|\lambda I - C| = 0$$

$$\left(\lambda - \frac{35}{12}\right)\cdot\left(\lambda - \frac{19}{12}\right) - \frac{7}{4}\cdot\frac{7}{4} = \lambda^2 - \frac{9}{2}\lambda + \frac{665}{144} = 0$$

$$\lambda_1 = +\frac{35}{12} \ ; \ \lambda_2 = +\frac{19}{12}$$

- for $\lambda_1 = +\frac{35}{12}$

$$\begin{bmatrix} \frac{35}{12}-\frac{35}{12} & +\frac{7}{4} \\ +\frac{7}{4} & \frac{35}{12}-\frac{19}{12} \end{bmatrix}\begin{bmatrix} 1 \\ \mu_2 \end{bmatrix} = 0 \Leftrightarrow$$

$$\begin{bmatrix} 0 & +\frac{7}{4} \\ +\frac{7}{4} & \frac{16}{12} \end{bmatrix}\begin{bmatrix} 1 \\ \mu_2 \end{bmatrix} = 0 \Leftrightarrow$$

$$\begin{bmatrix} 0 + \frac{7}{4}\mu_2 \\ \frac{7}{4} + \frac{16}{12}\mu_2 \end{bmatrix} = 0 \Leftrightarrow \begin{bmatrix} 0 \\ \frac{7}{4} \end{bmatrix} + \begin{bmatrix} \frac{7}{4}\mu_2 \\ \frac{16}{12}\mu_2 \end{bmatrix} = 0 \Leftrightarrow$$

$$\begin{bmatrix} 0 \\ \frac{7}{4} \end{bmatrix} = \begin{bmatrix} -\frac{7}{4}\mu_2 \\ -\frac{16}{12}\mu_2 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 0 \\ \frac{7}{4} \end{bmatrix} = \begin{bmatrix} -\frac{7}{4} \\ -\frac{16}{12} \end{bmatrix}\mu_2 \Leftrightarrow$$

[illegible]

$$\Leftrightarrow \mu_2 = \frac{-21}{37}$$

Normalized

$$U_1 = \begin{pmatrix} 1 \\ -\frac{21}{37} \end{pmatrix} \approx \begin{pmatrix} 1 \\ -0{,}567567 \end{pmatrix} \Rightarrow \begin{pmatrix} 0{,}869685 \\ -0{,}493607 \end{pmatrix}$$

- for $\lambda_2 = +\frac{19}{12}$

$$\begin{bmatrix} \frac{19}{12}-\frac{35}{12} & +\frac{7}{4} \\ +\frac{7}{4} & \frac{19}{12}-\frac{19}{12} \end{bmatrix}\begin{bmatrix} 1 \\ \mu_2 \end{bmatrix} = 0 \Leftrightarrow$$

$$\begin{bmatrix} -\frac{16}{12} & +\frac{7}{4} \\ +\frac{7}{4} & 0 \end{bmatrix}\begin{bmatrix} 1 \\ \mu_2 \end{bmatrix} = 0 \Leftrightarrow$$

$$\begin{bmatrix} -\frac{16}{12} + \frac{7}{4}\mu_2 \\ \frac{7}{4} + 0\,\mu_2 \end{bmatrix} = 0 \Leftrightarrow \begin{bmatrix} -\frac{16}{12} \\ \frac{7}{4} \end{bmatrix} + \begin{bmatrix} \frac{7}{4}\mu_2 \\ 0\,\mu_2 \end{bmatrix} = 0 \Leftrightarrow$$

$$\begin{bmatrix} -\frac{16}{12} \\ \frac{7}{4} \end{bmatrix} = \begin{bmatrix} -\frac{7}{4}\mu_2 \\ 0\,\mu_2 \end{bmatrix} \Leftrightarrow \begin{bmatrix} -\frac{16}{12} \\ \frac{7}{4} \end{bmatrix} = \begin{bmatrix} -\frac{7}{4} \\ 0 \end{bmatrix}\mu_2 \Leftrightarrow$$

[illegible]

$$\Leftrightarrow \mu_2 = -\frac{5}{21}$$

Normalized

$$U_1 = \begin{pmatrix} 1 \\ -\frac{5}{21} \end{pmatrix} \approx \begin{pmatrix} 1 \\ -0{,}238095 \end{pmatrix} \Rightarrow \begin{pmatrix} 0{,}972806 \\ -0{,}23162 \end{pmatrix}$$

$$U = \begin{pmatrix} 0{,}869685 & 0{,}972806 \\ -0{,}493607 & -0{,}23162 \end{pmatrix}$$

b) The most significant is U_1 because $\lambda_1 > \lambda_2$

($\lambda_1 < 1$ the Kaiser criterion is valid)

$$W = \begin{pmatrix} 0{,}869685 \\ -0{,}493607 \end{pmatrix} \qquad Z = W^T x$$

$$Z_1 = (0{,}869685 \ , -0{,}493607)\cdot\begin{pmatrix} 2 \\ 4 \end{pmatrix} = -0{,}235058$$

$$Z_2 = (0{,}869685 \ , -0{,}493607)\cdot\begin{pmatrix} 4 \\ 2 \end{pmatrix} = 2.491526$$

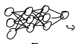
$$Z_3 = (0{,}869685 \ , -0{,}493607)\cdot\begin{pmatrix} 0 \\ 4 \end{pmatrix} = -1{,}974428$$

$$Z_4 = (0{,}869685 \ , -0{,}493607)\cdot\begin{pmatrix} 1 \\ 5 \end{pmatrix} = -1{,}59835$$

IV VC Dimension

a)

1) MLP $\Rightarrow$ 3 3 3 1 $\Rightarrow$



$\hookrightarrow$ we have 3 inputs and 1 output

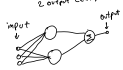
- input $\rightarrow$ first hidden layer
  - 3×3 weight matrix $\Rightarrow$ 9 parameters
  - 3×1 bias vector $\Rightarrow$ 3 parameters
- first hidden layer $\rightarrow$ second hidden layer
  - 3×3 weight matrix $\Rightarrow$ 9 parameters
  - 3×1 bias vector $\Rightarrow$ 3 parameters
- second hidden layer $\rightarrow$ output
  - 1×3 weight matrix $\Rightarrow$ 3 parameters
  - 1×1 bias vector $\Rightarrow$ 1 parameter

$\Rightarrow 9+3+9+3+3+1 = \underline{28}$

2) RBF Network with 2 hidden units with σ equal + logistic regression for output

3D input
2 output elements

$$F(x) = \sum_{k=1}^{K} w_k \cdot \phi\left(\|x - c_k\|\right)$$



taking into consideration that the radial basis function does not have a β parameter.

1×1 Value of Sigma for the radial basis function = 1

2×1 weight matrix between hidden layer and output = 2

1×1 bias vector = 1

$1+2+1 = 4$

The discribed RBF Network has the smallest VC Dimension, compared to the MLP.

b) The relation between regularization and VC Dimension is inverse. Regularization forbids some hypothesis (by constraining the learning to simpler hypothesis) and thus, reducing the VC dimension.
This improves generalization.