

CPE 721 – 2012 - RNs Feedforward

2ª Série de Exercícios - Pré-processamento

1 – Considere as variáveis x_1, x_2 e y , todas com média nula e desvio padrão σ_1, σ_2 e σ_y , respectivamente. r_{12} é a correlação (coeficiente de correlação de Pearson) entre x_1 e x_2 , r_{1y} é a correlação entre x_1 e y , e r_{2y} é a correlação entre x_2 e y . A variável $z = x_2 - ax_1$ é a variável “descorrelacionada” da variável x_1 , e r_{zy} é a correlação entre z e y . O parâmetro a é obtido pela minimização de $E(x_2 - ax_1)^2$. Prove que:

$$a = \frac{\sigma_2}{\sigma_1} r_{12} \quad \text{e} \quad r_{zy} = \frac{r_{2y} - r_{1y} r_{12}}{\sqrt{1 - r_{12}^2}}$$

Handwritten derivation:

$$\mu_x = 0 \quad \sigma_x^2 = E x^2$$

$$r_{xy} = \frac{E xy}{\sigma_x \sigma_y} \quad \text{coef. correlação}$$

$$E (x_2 - a x_1)^2 =$$

$$= E x_2^2 + a^2 E x_1^2 - 2a E x_1 x_2$$

$$\frac{dE}{da} = 2a E x_1^2 - 2 E x_1 x_2 = 0$$

$$a = \frac{E x_1 x_2}{E x_1^2} = \frac{r_{12} \sigma_1 \sigma_2}{\sigma_1^2} = \frac{\sigma_2}{\sigma_1} r_{12}$$

$$r_{2y} = \frac{E z y}{\sigma_2 \sigma_y} = \frac{E (x_2 - a x_1) y}{(E (x_2 - a x_1)^2 E y^2)^{1/2}}$$

$$E z y = E x_2 y - a E x_1 y$$

$$= \sigma_2 \sigma_y r_{2y} - \frac{\sigma_2}{\cancel{\sigma_1}} r_{12} \cancel{\sigma_1} \sigma_y r_{1y}$$

$$E z^2 = E x_2^2 + a^2 E x_1^2 - 2a E x_1 x_2$$

$$= \sigma_2^2 + \frac{\sigma_2^2}{\cancel{\sigma_1^2}} r_{12}^2 \cancel{\sigma_1^2} - 2 \frac{\sigma_2}{\cancel{\sigma_1}} r_{12} \cancel{\sigma_1} \sigma_2 r_{12}$$

$$r_{2y} = \frac{r_{2y} - r_{12} r_{1y}}{\sqrt{1 - r_{12}^2}}$$

2 – Apresente a normalização a ser usada para os pares entrada-saída (X, Y) da Tabela 1 abaixo.

$$x_i = \frac{1}{\sigma_{x_i}} (X_i - \mu_{x_i}) \quad >>>$$

$$\mu_x = 0 \text{ e } \sigma_x = 1$$

Variável	X1	X2	X3	X4	Y
Média	9,661	-0,978	2,087	4,916	3,438
D.P.	2,087	1,082	0,611	1,402	1,845

2.1 - Use correlação para verificar se alguma variável de entrada pode ser eliminada.

Verificação de entradas relevantes e interdependência das entradas

$$N = 36 \quad \frac{2}{\sqrt{N}} = 0,33$$

Correlação	X1	X2	X3	X4	Y
X1	1,00	-0,020	-0,066	-0,598	-0,588
X2		1,00	-0,111	-0,072	0,081
X3			1,00	0,292	-0,414
X4				1,00	0,570

Excluir x_2

Excluir x_4 ?

Descorrelacionando x_4 de x_1

$$r_{4'y} = \frac{r_{4y} - r_{1y}r_{14}}{\sqrt{1 - r_{14}^2}} = \frac{(0,570) - (-0,588)(-0,598)}{\sqrt{1 - (-0,598)^2}} = 0,272 < 0,33$$

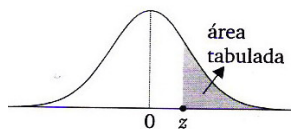
Excluir x_4 ? possivelmente

2.2 - Verifique a presença de possíveis intrusos. Esta verificação é mais fácil após a normalização (mas também é possível com os dados brutos)

Verificação de possíveis intrusos

$1/36 \sim 2,8 \%$

$1,4 \% = 0,014 = \text{erf}(2,20)$



Valores esperados

$-2,2 < x < 2,2$

z	Segunda decimal de z									
	0	1	2	3	4	5	6	7	8	9
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064

Valores normalizados $\mu = 0 \sigma = 1$

x1	x2	x3	x4	y	x1	x2	x3	x4	y
-0,925	-0,547	-1,156	-0,253	0,640	0,895	0,857	0,807	-0,645	-0,974
-0,225	-0,759	1,232	-0,239	-0,671	-0,551	1,633	-0,190	-0,004	0,754
-0,163	-0,251	0,888	1,222	0,028	0,267	-1,850	0,905	0,0457	0,537
-1,083	0,312	-1,008	0,908	0,922	-0,110	-0,269	-0,043	-0,360	0,190
1,355	0,839	-1,630	-1,415	-0,107	0,608	-0,297	-0,223	-0,118	-0,508
-0,709	-0,953	0,627	0,680	0,418	0,028	-0,962	0,365	-0,702	-0,660
0,042	-1,406	-0,992	0,045	0,667	1,307	-0,316	-0,518	0,295	-0,557
-0,263	-1,785	1,379	0,765	-0,280	0,152	-1,609	-0,190	0,937	0,548
0,028	-0,731	-1,237	-1,102	0,082	-1,097	1,365	-0,730	-0,011	0,841
-1,246	0,414	0,398	0,159	0,429	1,633	-0,297	-1,270	-2,392	-1,451
-2,429	-0,288	1,068	0,288	-0,156	0,435	1,033	1,150	0,259	-1,002
-0,254	0,626	1,199	1,364	0,125	-0,354	-1,323	-0,485	1,044	0,456
0,459	0,645	-1,695	0,651	2,998	1,245	-0,657	0,643	-1,237	-1,402
1,647	-0,279	1,412	-1,194	-2,307	-1,040	0,682	0,872	1,578	0,694
1,791	-0,205	-1,335	-1,993	-1,348	-0,283	0,811	0,757	1,771	0,862
-1,346	-0,639	0,103	1,528	0,667	-1,327	0,931	-1,450	-0,018	1,231
1,652	1,107	1,117	-0,617	-1,706	-0,450	1,144	-1,139	0,302	0,683
-0,033	1,255	0,986	-0,310	-0,698	0,344	1,772	-0,616	-1,230	0,055

3 – Projete uma rede neural para realizar o mapeamento da Tabela 1 acima. Aloque 20 pares para treinamento, 8 para teste e 8 para validação. Os possíveis intrusos, se existirem, devem ser alocados no conjunto de teste, e os demais pares sorteados pelos três conjuntos. Projete a rede usando apenas as variáveis que persistiram após a eliminação usando correlação linear. Não é necessário fazer a validação cruzada, mas faça o acompanhamento durante o treinamento e a crítica pós treinamento. Sugestão: utilize uma rede com duas camadas de neurônios, com 4 neurônios na camada intermediária.

3.1 - Repita inserindo uma a uma as variáveis eliminadas pelo critério de correlação linear. Compare os resultados obtidos.