Trabalho 4
PCA e Classificador SVM
Disciplina de Reconhecimento de Padrões
Hugo Silveira Sousa - 378998

2.1) Apresentação da base de dados.

A base de dados “Divorce Predictors data set” será utilizada para implementação dos algoritmos do Trabalho.

A base tem 170 amostras, esses dados apresentam as informações divididas em 2 classes, de pessoas em casamentos felizes (rótulo 0, 86 amostras, 51%) e pessoas divorciadas (rótulo 1, 84 amostras, 49%), a base foi coletada a partir de entrevistas com participantes de várias regiões da Turquia, onde 84 participantes são homens e 86 são mulheres.

A base tem 54 atributos, que são 54 perguntas sobre o casamento do participante, e são respondidas em uma escala de 5 valores (Divorce Predictors Scale), onde:

0 = Nunca | 1 = Raramente | 2 = Média | 3 = Frequentemente | 4 = Sempre

O objetivo é a partir de uma instância classificar o participante em pertecente a um casamento feliz ou divorciado.

Vale ressaltar que os resultados dos estudos dessa base podem variar quando testados com informações de participantes de outros países, pois aspectos culturais variam entre populações, nessa base 74 participantes se casaram por amor, e 96 se casaram por casamentos arranjados, que não é tão comum em nossa cultura, por exemplo.

Artigo sobre a base:
Yöntem, Mustafa Kemal, et al. "Divorce prediction using correlation based feature selection and artificial neural networks." *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi* 9.1 (2019): 259-273.
Fonte da base:

- https://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set
- https://www.kaggle.com/andrewmvd/divorce-prediction

Perguntas:

1. If one of us apologizes when our discussion deteriorates, the discussion ends.
2. I know we can ignore our differences, even if things get hard sometimes.
3. When we need it, we can take our discussions with my spouse from the beginning and correct it.
4. When I discuss with my spouse, to contact him will eventually work.
5. The time I spent with my wife is special for us.
6. We don't have time at home as partners.
7. We are like two strangers who share the same environment at home rather than family.
8. I enjoy our holidays with my wife.
9. I enjoy traveling with my wife.
10. Most of our goals are common to my spouse.
11. I think that one day in the future, when I look back, I see that my spouse and I have been in harmony with each other.
12. My spouse and I have similar values in terms of personal freedom.
13. My spouse and I have similar sense of entertainment.
14. Most of our goals for people (children, friends, etc.) are the same.
15. Our dreams with my spouse are similar and harmonious.
16. We're compatible with my spouse about what love should be.
17. We share the same views about being happy in our life with my spouse
18. My spouse and I have similar ideas about how marriage should be

19. My spouse and I have similar ideas about how roles should be in marriage
20. My spouse and I have similar values in trust.
21. I know exactly what my wife likes.
22. I know how my spouse wants to be taken care of when she/he sick.
23. I know my spouse's favorite food.
24. I can tell you what kind of stress my spouse is facing in her/his life.
25. I have knowledge of my spouse's inner world.
26. I know my spouse's basic anxieties.
27. I know what my spouse's current sources of stress are.
28. I know my spouse's hopes and wishes.
29. I know my spouse very well.
30. I know my spouse's friends and their social relationships.
31. I feel aggressive when I argue with my spouse.
32. When discussing with my spouse, I usually use expressions such as ‘you always’ or ‘you never’ .
33. I can use negative statements about my spouse's personality during our discussions.
34. I can use offensive expressions during our discussions.
35. I can insult my spouse during our discussions.
36. I can be humiliating when we discussions.
37. My discussion with my spouse is not calm.
38. I hate my spouse's way of open a subject.
39. Our discussions often occur suddenly.
40. We're just starting a discussion before I know what's going on.
41. When I talk to my spouse about something, my calm suddenly breaks.
42. When I argue with my spouse, ı only go out and I don't say a word.
43. I mostly stay silent to calm the environment a little bit.
44. Sometimes I think it's good for me to leave home for a while.
45. I'd rather stay silent than discuss with my spouse.
46. Even if I'm right in the discussion, I stay silent to hurt my spouse.
47. When I discuss with my spouse, I stay silent because I am afraid of not being able to control my anger.
48. I feel right in our discussions.
49. I have nothing to do with what I've been accused of.
50. I'm not actually the one who's guilty about what I'm accused of.
51. I'm not the one who's wrong about problems at home.
52. I wouldn't hesitate to tell my spouse about her/his inadequacy.
53. When I discuss, I remind my spouse of her/his inadequacy.
54. I'm not afraid to tell my spouse about her/his incompetence.

2.2) PCA

Arquivos:

Func/calc_matriz_projecao_PCA.m

Func/PCA.m

Rth	Número de Componentes
0.8	1
0.9	**7**
0.99	31
0.999	45
0.9999	52

- Para os experimentos foi escolhido o Rth=0.9, com 7 componentes

2.3) SVM
Arquivos:
Func/SVM_treino.m
Func/SVM_teste.m

Resultados:
Arquivo:
experimentos.m

Condições dos experimentos:

- 20 iterações para cada experimento;
- A cada iteração a base é permutada aleatoriamente;
- A semente da função rand foi fixada em 0, para ter os mesmos resultados;
- O PCA usa o Rth = 0.9;
- Valores para C = 0.01, 0.1, 1, 5, 10, 100;
- Valores para gamma = 0.01, 0.1, 1, 5, 10, 100;
- Dois kernel's utilizados, o RBF e o Linear;

Obs: As matrizes de confusão foram geradas para cada iteração de cada experimento, no seguinte formato:

TP	**FP**
FN	**TN**

Experimentos para o kernel RBF:

Nº	Condições do Experimento		(em %)	--
	C	gamma	Média das Acurácias	STD das Acurácias
1	0.01	0.01	97.352941	2.403566
2	0.01	0.1	97.647059	1.865594
3	0.01	1	59.803922	21.055920
4	0.01	5	49.313725	9.527050
5	0.01	10	48.921569	9.163279
6	0.01	100	48.921569	9.163279
7	0.1	0.01	97.647059	1.865594
8	0.1	0.1	97.647059	1.865594
9	0.1	1	85.490196	6.015016
10	0.1	5	59.215686	10.484142
11	0.1	10	59.019608	10.354020

12	0.1	100	59.019608	10.354020
13	1	0.01	97.647059	1.865594
14	1	0.1	97.647059	1.865594
15	1	1	92.058824	5.225640
16	1	5	63.235294	8.719720
17	1	10	62.156863	9.177070
18	1	100	62.156863	9.177070
19	5	0.01	97.647059	1.865594
20	5	0.1	98.235294	1.788058
21	5	1	92.254902	4.864667
22	5	5	63.431373	8.798267
23	5	10	62.156863	9.177070
24	5	100	62.156863	9.177070
25	10	0.01	97.647059	1.865594
26	10	0.1	98.235294	1.788058
27	10	1	92.254902	4.864667
28	10	5	63.431373	8.798267
29	10	10	62.156863	9.177070
30	10	100	62.156863	9.177070
31	**100**	**0.01**	**98.431373**	**1.865594**
32	100	0.1	98.235294	1.788058
33	100	1	92.254902	4.864667
34	100	5	63.431373	8.798267
35	100	10	62.156863	9.177070
36	100	100	62.156863	9.177070

Os experimentos(RBF) com taxas médias de acurácia maiores que 98% estão em destaque.

Experimentos para o kernel Linear:

Nº	Condições do Experimento	(em %)	--
	C	Média das Acurácias	STD das Acurácias
37	0.01	97.941176	1.739296
38	0.1	99.019608	1.491933
39	**1**	**99.215686**	**1.478308**
40	**5**	**99.215686**	**1.478308**
41	**10**	**99.215686**	**1.478308**
42	**100**	**99.215686**	**1.478308**

Os experimentos(Linear) com taxas médias de acurácia maiores que 99% estão em destaque.

Os experimentos **31** foi o melhor experimento com o **Kernel RBF**, com **C = 100** e **gamma = 0.01**, obtiveram a **média das acurácias = 98.431373**, com **desvio padrão = 1.865594.**

Os experimentos **39, 40, 41 e 42** foram os melhores experimentos com o **Kernel Linear**, com **C = 1, 5, 10 e 100,** obtiveram a **média das acurácias = 99.215686**, com **desvio padrão = 1.478308**.

Os experimentos **39, 40, 41 e 42** foram os melhores experimentos.

Para realizar testes use o **Arquivo: main.m**