

Aprendizaje automático
Departamento de Ingeniería en Informática
ITBA

Trabajo Práctico 6

Objetivo: Aplicar diversos métodos de estadística no paramétrica para aprender a hacer inferencia a partir de datos experimentales

Aprendizaje basado en instancias

1. Usando el clasificador basado en los k-vecinos más cercanos separar a los “datos de los lirios Fisher” (ver ej. 5 tp2) en 3 grupos:
 - (a) considerando el ancho y largo de los pétalos.
 - (b) considerando el ancho y largo de los sépalos.
 - (c) considerando las cuatro variables.
 - i. Utilizando todos los datos como conjunto de entrenamiento y de testeo.
 - ii. Separando 100 datos al azar para usarlos como conjunto de entrenamiento y usando los 50 datos restantes como conjunto de testeo.
 - (d) Calcular el porcentaje de datos correctamente clasificados para los ítems anteriores.
 - (e) Comparar los resultados con los obtenidos en el ejercicio 5 del Tp2, en el ejercicio 2 del Tp4 y en ejercicio 3 del Tp5.
2. Los siguientes datos corresponden a la velocidad (en millas por hora) y la distancia que requieren para frenar (en pies) de 50 automóviles.
Los datos fueron registrados en 1920.



automóvil	velocidad	distancia	automóvil	velocidad	distancia	automóvil	velocidad	distancia
1	4	2	18	13	34	35	18	84
2	4	10	19	13	46	36	19	36
3	7	4	20	14	26	37	19	46
4	7	22	21	14	36	38	19	68
5	8	16	22	14	60	39	20	32
6	9	10	23	14	80	40	20	48
7	10	18	24	15	20	41	20	52
8	10	26	25	15	26	42	20	56
9	10	34	26	15	54	43	20	64
10	11	17	27	16	32	44	22	66
11	11	28	28	16	40	45	23	54

12	12	14	29	17	32	46	24	70
13	12	20	30	17	40	47	24	92
14	12	24	31	17	50	48	24	93
15	12	28	32	18	42	49	24	120
16	13	26	33	18	56	50	25	85
17	13	34	34	18	76			

- (a) Ajustar los datos mediante una regresión lineal local pesada con el núcleo gaussiano.
- (b) Ajustar los datos mediante una regresión lineal local pesada con el núcleo tricubo.
- Considerando 5 vecinos.
 - Considerando 10 vecinos.
 - Considerando 30 vecinos.
- ¿Qué observa en los ajustes cuando aumenta el número de vecinos?
3. Se realizó un experimento para estudiar si la colostiramina era una droga efectiva para bajar los niveles de colesterol en sangre. Los pacientes debían tomar 6 paquetes por día, pero algunos no siguieron las indicaciones y tomaron una dosis menor.
- Se midieron 2 variables: x = porcentaje de acatamiento e y = disminución en el nivel de colesterol

n	x	y	n	x	y	n	x	y	n	x	y
1	0	-5.25	42	34	1.00	83	87	11.50	124	99	82.75
2	27	-1.50	43	78	36.00	84	98	54.75	125	20	39.00
3	71	59.50	44	96	28.75	85	11	-13.00	126	57	13.75
4	95	32.50	45	7	-10.50	86	47	32.50	127	93	71.00
5	0	-7.25	46	34	7.75	87	87	2.75	128	99	85.00
6	28	23.50	47	78	39.50	88	98	26.75	129	21	-5.75
7	71	14.75	48	96	26.75	89	12	24.00	130	57	48.75
8	95	70.75	49	8	19.75	90	50	54.50	131	93	37.75
9	0	-6.25	50	35	-15.75	91	88	48.75	132	99	70.00
10	29	33.00	51	81	1.00	92	98	80.00	133	21	-21.00
11	72	63.00	52	96	56.00	93	13	2.50	134	58	43.00
12	95	18.25	53	8	-0.50	94	50	-4.25	135	93	41.00
13	0	11.50	54	36	33.50	95	89	56.75	136	100	92.00
14	31	4.25	55	82	53.50	96	98	42.25	137	21	0.25
15	72	0.00	56	96	47.50	97	15	3.00	138	60	27.75
16	95	76.00	57	8	29.25	98	51	42.75	139	93	9.75
17	2	21.00	58	36	36.25	99	90	29.25	140	100	73.75
18	32	18.75	59	84	46.50	100	98	6.00	141	22	-10.25
19	73	42.00	60	96	30.25	101	15	5.50	142	62	44.50
20	95	75.75	61	8	36.25	102	51	62.75	143	93	53.75
21	2	-23.00	62	37	5.50	103	90	72.50	144	100	54.00
22	32	8.50	63	85	51.00	104	98	104.75	145	24	-0.50
23	74	41.25	64	96	21.00	105	16	21.25	146	64	22.50

24	95	78.75	65	9	10.75	106	52	64.25	147	94	62.50
25	2	5.75	66	38	25.50	107	91	41.75	148	100	69.50
26	33	3.25	67	85	39.00	108	98	94.25	149	25	-19.00
27	75	36.25	68	97	79.00	109	16	29.75	150	64	-14.50
28	95	54.75	69	9	19.50	110	53	30.25	151	94	39.00
29	3	3.25	70	41	20.25	111	92	48.50	152	100	101.50
30	33	27.75	71	87	-0.25	112	98	41.25	153	25	15.75
31	76	66.50	72	97	69.00	113	17	7.50	154	64	-20.75
32	95	77.00	73	9	17.25	114	54	14.75	155	94	3.25
33	3	8.75	74	43	33.25	115	92	61.25	156	100	68.00
34	34	30.75	75	87	1.00	116	98	40.25	157	26	6.00
35	77	61.75	76	97	80.00	117	18	-16.50	158	67	46.25
36	96	68.00	77	10	3.50	118	54	47.25	159	94	60.00
37	4	8.25	78	45	56.75	119	92	29.50	160	100	44.75
38	34	-1.50	79	87	46.75	120	99	51.50	161	27	10.50
39	77	14.00	80	97	86.00	121	20	4.50	162	68	39.50
40	96	73.00	81	10	11.25	122	56	18.00	163	95	113.25
41	4	-10.25	82	45	4.25	123	92	59.75	164	100	86.75

- (a) Graficar ‘porcentaje de acatamiento’ vs. ‘disminución en el nivel de colesterol’.
- (b) Ajustar los datos mediante un polinomio adecuado.
- (c) Ajustar los datos mediante una regresión lineal local pesada.
- (d) Analizar los resultados obtenidos en los ítems anteriores.