Aprendizagem 2024/25

Homework II – Group 99

(ist1106059, ist1106329)

I. Pen-and-paper

	y_1	y_2	y_{out}
x_1	A	0	P
x_2	B	1	P
x_3	A	1	P
x_4	A	0	P
x_5	B	0	N
x_6	B	0	N
x_7	A	1	N
x_8	B	1	N

1) Calcular Hamming Distance:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	-	2	1	0	1	1	1	2
			P	P	N	N	N	

$\widehat{z_1} = mode(P, P, N, N, N) = \boldsymbol{N}$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_2	2	-	1	2	1	1	1	0
			P		N	N	N	N

$\widehat{z_2} = mode(P, N, N, N, N) = \boldsymbol{N}$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_3	1	1	-	1	2	2	0	1
	P	P		P			N	N

$\widehat{z_3} = mode(P, P, P, N, N) = \boldsymbol{P}$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_4	0	2	1	-	1	1	1	2
	P		P		N	N	N	

$\widehat{z_4} = mode(P, P, N, N, N) = \boldsymbol{N}$

Homework II – Group 99

(ist1106059, ist1106329)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_5	1	1	2	1	-	0	2	1
	P	P		P		N		N

$$\hat{z}_5 = mode(P, P, P, N, N) = \boldsymbol{P}$$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_6	1	1	2	1	0	-	2	1
	P	P		P	N			N

$$\hat{z}_6 = mode(P, P, P, N, N) = \boldsymbol{P}$$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_7	1	1	0	1	2	2	-	1
	P	P	P	P				N

$$\hat{z}_7 = mode(P, P, P, P, N) = \boldsymbol{P}$$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_8	2	0	1	2	1	1	1	-
		P	P		N	N	N	

$$\hat{z}_8 = mode(P, P, N, N, N) = \boldsymbol{N}$$

Confusing Matrix e F1-measure:

Predicted \ Real	P	N
P	1	3
N	3	1

$$recall = \frac{TP}{TP + FN} = \frac{1}{1+3} = \frac{1}{4}$$

$$precision = \frac{TP}{TP + FP} = \frac{1}{1+3} = \frac{1}{4}$$

$$F1 = \frac{2}{4+4} = \boldsymbol{\frac{1}{4}}$$

2) Ao observar o data set, conseguimos perceber que a variável y_1 é mais relevante que y_2 uma vez que existe uma menor variação dos valores de y_1 em cada valor da variável target, ao contrário dos valores de y_2 que variam muito mais. Posto isto, podemos calcular a distância de Hamming tendo em conta apenas a variável y_1:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	-	1	0	0	1	1	0	1

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_5	1	0	1	1	-	0	1	0

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_2	1	-	1	1	0	0	1	0

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_6	1	0	1	1	0	-	1	0

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_3	0	1	-	0	1	1	0	1

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_7	0	1	0	0	1	1	-	1

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_4	0	1	0	-	1	1	0	1

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_8	1	0	1	1	0	0	1	-

De acordo com as distâncias de Hamming, vemos que o melhor valor para **k** é 3, que corresponde ao menor valor de k que não gera empates nas distâncias.

Posto isto temos:

$$\hat{z_1} = mode(P, P, N) = \boldsymbol{P}$$
$$\hat{z_2} = mode(N, N, N) = \boldsymbol{N}$$
$$\hat{z_3} = mode(P, P, N) = \boldsymbol{P}$$
$$\hat{z_4} = mode(P, P, N) = \boldsymbol{P}$$
$$\hat{z_5} = mode(P, N, N) = \boldsymbol{N}$$
$$\hat{z_6} = mode(P, N, N) = \boldsymbol{N}$$
$$\hat{z_7} = mode(P, P, P) = \boldsymbol{P}$$
$$\hat{z_8} = mode(P, N, N) = \boldsymbol{N}$$

Calculando a F1-measure como no exercício anterior obtemos:

$$recall = \frac{TP}{TP + FN} = \frac{3}{3+1} = \frac{3}{4} \qquad precision = \frac{TP}{TP + FP} = \frac{3}{3+1} = \frac{3}{4}$$

$$F1 = \frac{2}{\frac{4}{3} + \frac{4}{3}} = \boldsymbol{\frac{3}{4}} = \boldsymbol{3} \times \boldsymbol{\frac{1}{4}}$$

Logo, com esta nova métrica conseguimos melhorar a performance em 3 vezes.

3)

	y_1	y_2	y_3	y_{out}
x_1	A	0	1.1	P
x_2	B	1	0.8	P
x_3	A	1	0.5	P
x_4	A	0	0.9	P
x_5	B	0	1.0	N
x_6	B	0	0.9	N
x_7	A	1	1.2	N
x_8	B	1	0.9	N
x_9	B	0	0.8	P

Priori:

$$P(class = P) = \frac{5}{9} \qquad P(class = N) = 1 - P(class = P) = \frac{4}{9}$$

Cálculo das distribuições Gaussianas:

$$P(y_3 = a) = N(y_3 \mid \mu, \sigma^2) \ = N(y_3 \mid 0.9, 0.04)$$

$$\mu = \frac{1.1 + 0.8 + 0.5 + 0.9 + 1 + 0.9 + 1.2 + 0.9 + 0.8}{9} = 0.9$$

$$\sigma^2 = \frac{(1.1-0.9)^2+(0.8-0.9)^2+(0.5-0.9)^2+(0.9-0.9)^2+(1.0-0.9)^2+(0.9-0.9)^2+(1.2-0.9)^2+(0.9-0.9)^2+(0.8-0.9)^2}{8} = 0.04$$

$$P(y_3 = a \mid class = P) = N(y_3 \mid class = P \mid \mu,\ \sigma^2) \ = N(y_3 \mid class = P \mid 0.82,\ 0.047)$$

$$\mu = \frac{1.1 + 0.8 + 0.5 + 0.9 + 0.8}{5} = 0.82$$

$$\sigma^2 = \frac{(1.1 - 0.82)^2 + (0.8 - 0.82)^2 + (0.5 - 0.82)^2 + (0.9 - 0.82)^2 + (0.8 - 0.82)^2}{4} = 0.047$$

$$P(y_3 = a \mid class = N) = N(y_3 \mid class = N \mid \mu,\ \sigma^2)\ = N(y_3 \mid class = N \mid 1,\ 0.02)$$

$$\mu = \frac{1 + 0.9 + 1.2 + 0.9}{4} = 1$$

$$\sigma^2 = \frac{(1-1)^2 + (0.9-1)^2 + (1.2-1)^2 + (0.9-1)^2}{3} = 0.02$$

Posteriori para todas as combinações de valores possíveis:

$$P(class = P \mid\ y_1 = A, y_2 = 0, y_3 = a) = \frac{P(y_1 = A, y_2 = 0, y_3 = a \mid class = P)P(class = P)}{P(y_1 = A, y_2 = 0, y_3 = a)}$$
$$= \frac{P(y_1 = A, y_2 = 0 \mid class = P)P(y_3 = a \mid class = P)P(class = P)}{P(y_1 = A, y_2 = 0)P(y_3 = a)}$$
$$= \frac{\frac{2}{5}\ N(y_3 \mid class = P \mid 0.82, 0.047)\frac{5}{9}}{\frac{2}{9}\ N(y_3 \mid 0.9, 0.04)} = \frac{N(y_3 \mid 0.82,0.047)}{N(y_3 \mid 0.9,0.04)}$$

$$P(class = P \mid\ y_1 = A, y_2 = 1, y_3 = a)$$
$$= \frac{P(y_1 = A, y_2 = 1 \mid class = P)P(y_3 = a \mid class = P)P(class = P)}{P(y_1 = A, y_2 = 1)P(y_3 = a)}$$
$$= \frac{\frac{1}{5}\ N(y_3 \mid class = P \mid 0.82, 0.047)\frac{5}{9}}{\frac{2}{9}\ N(y_3 \mid 0.9, 0.04)} = \frac{N(y_3 \mid 0.82,0.047)}{2 \times N(y_3 \mid 0.9,0.04)}$$

$$P(class = P \mid\ y_1 = B, y_2 = 0, y_3 = a)$$
$$= \frac{P(y_1 = B, y_2 = 0 \mid class = P)P(y_3 = a \mid class = P)P(class = P)}{P(y_1 = B, y_2 = 0)P(y_3 = a)}$$
$$= \frac{\frac{1}{5}\ N(y_3 \mid class = P \mid 0.82, 0.047)\frac{5}{9}}{\frac{3}{9}\ N(y_3 \mid 0.9, 0.04)} = \frac{N(y_3 \mid 0.82,0.047)}{3 \times N(y_3 \mid 0.9,0.04)}$$

$$P(class = P \mid\ y_1 = B, y_2 = 1, y_3 = a)$$
$$= \frac{P(y_1 = B, y_2 = 1 \mid class = P)P(y_3 = a \mid class = P)P(class = P)}{P(y_1 = B, y_2 = 1)P(y_3 = a)}$$
$$= \frac{\frac{1}{5}\ N(y_3 \mid class = P \mid 0.82, 0.047)\frac{5}{9}}{\frac{2}{9}\ N(y_3 \mid 0.9, 0.04)} = \frac{N(y_3 \mid 0.82,0.047)}{2 \times N(y_3 \mid 0.9,0.04)}$$

$P(class = N \mid\ y_1 = A, y_2 = 0, y_3 = a)$

$$= \frac{P(y_1 = A, y_2 = 0 \mid class = N)P(y_3 = a \mid class = \ N)P(class = N)}{P(y_1 = \ A, y_2 = 0)P(y_3 = a)}$$

$$= \frac{0\ N(y_3 \mid class = \ N \mid 1.0,\ 0.02)\frac{4}{9}}{\frac{2}{9}\ N(y_3 \mid 0.9, 0.04)} = 0$$

$P(class = N \mid\ y_1 = A, y_2 = 1, y_3 = a)$

$$= \frac{P(y_1 = A, y_2 = 1 \mid class = N)P(y_3 = a \mid class = \ N)P(class = N)}{P(y_1 = \ A, y_2 = 1)P(y_3 = a)}$$

$$= \frac{\frac{1}{4}\ N(y_3 \mid class = \ N \mid 1.0,\ 0.02)\frac{4}{9}}{\frac{2}{9}\ N(y_3 \mid 0.9, 0.04)} = \frac{N(y_3 \mid 1,0.02)}{2 \times N(y_3 \mid 0.9,0.04)}$$

$P(class = N \mid\ y_1 = B, y_2 = 0, y_3 = a)$

$$= \frac{P(y_1 = B, y_2 = 0 \mid class = N)P(y_3 = a \mid class = \ N)P(class = N)}{P(y_1 = \ B, y_2 = 0)P(y_3 = a)}$$

$$= \frac{\frac{2}{4}\ N(y_3 \mid class = \ N \mid 1.0,\ 0.02)\frac{4}{9}}{\frac{3}{9}\ N(y_3 \mid 0.9, 0.04)} = \frac{2 \times N(y_3 \mid 1,0.02)}{3 \times N(y_3 \mid 0.9,0.04)}$$

$P(class = N \mid\ y_1 = B, y_2 = 1, y_3 = a)$

$$= \frac{P(y_1 = B, y_2 = 1 \mid class = N)P(y_3 = a \mid class = \ N)P(class = N)}{P(y_1 = \ B, y_2 = 1)P(y_3 = a)}$$

$$= \frac{\frac{1}{4}\ N(y_3 \mid class = \ N \mid 1,\ 0.02)\frac{4}{9}}{\frac{2}{9}\ N(y_3 \mid 0.9, 0.04)} = \frac{N(y_3 \mid 1, 0.02)}{2 \times N(y_3 \mid 0.9,0.04)}$$

4)

	y_1	y_2	y_3	y_{out}
x_1	A	0	1.1	P
...	...	...	...	...
x_{10}	A	1	0.8	?
x_{11}	B	1	1	?
x_{12}	B	0	0.9	?

$$P(class = P \mid y_1 = A, y_2 = 1, y_3 = 0.8) = \frac{N(y_3 = 0.8 \mid 0.82, 0.047)}{2 \times N(y_3 = 0.8 \mid 0.9, 0.04)} = \frac{1.8324}{2 \times N(y_3 = 0.8|0.9, 0.04)}$$

$$P(class = N \mid y_1 = A, y_2 = 1, y_3 = 0.8) = \frac{N(y_3 = 0.8|1, 0.02)}{2 \times N(y_3 = 0.8|0.9, 0.04)} = \frac{1.0378}{2 \times N(y_3 = 0.8|0.9, 0.04)}$$

$$Como\ P(class = P \mid y_1 = A, y_2 = 1, y_3 = 0.8) > P(class = N \mid y_1 = A, y_2 = 1, y_3 = 0.8), \widehat{z_{10}} = \boldsymbol{P}$$

$$P(class = P \mid y_1 = B, y_2 = 1, y_3 = 1) = \frac{N(y_3 = 1 \mid 0.82, 0.047)}{2 \times N(y_3 = 1 \mid 0.9, 0.04)} = \frac{1.3037}{2 \times N(y_3 = 0.1|0.9, 0.04)}$$

$$P(class = N \mid y_1 = B, y_2 = 1, y_3 = 1) = \frac{N(y_3 = 1 \mid 1, 0.02)}{2 \times N(y_3 = 1 \mid 0.9, 0.04)} = \frac{2.8209}{2 \times N(y_3 = 0.1|0.9, 0.04)}$$

$$Como\ P(class = P \mid y_1 = B, y_2 = 1, y_3 = 1) < P(class = N \mid y_1 = B, y_2 = 1, y_3 = 1),\ \widehat{z_{11}} = \boldsymbol{N}$$

$$P(class = P \mid y_1 = B, y_2 = 0, y_3 = 0.9) = \frac{N(y_3 = 0.9 \mid 0.82, 0.047)}{3 \times N(y_3 = 0.9 \mid 0.9, 0.04)} = \frac{1.7191}{3 \times N(y_3 = 0.9|0.9, 0.04)}$$

$$P(class = N \mid y_1 = B, y_2 = 0, y_3 = 0.9) = \frac{2 \times N(y_3 = 0.9 \mid 1, 0.02)}{3 \times N(y_3 = 0.9 \mid 0.9, 0.04)} = \frac{2 \times 2.1970}{3 \times N(y_3 = 0.9|0.9, 0.04)}$$

$$Como\ P(class = P \mid y_1 = B, y_2 = 0, y_3 = 0.9) < P(class = N \mid y_1 = B, y_2 = 0, y_3 = 0.9),\ \widehat{z_{12}} = \boldsymbol{N}$$

5)

	termo	classe
t_1	Amazing	P
t_2	run	P
t_3	I	P
t_4	like	P
t_5	it	P
t_6	Too	N
t_7	tired	N
t_8	Bad	N
t_9	run	N

$V = 8$
$N_N = 4$
$N_P = 5$

$$P(I \mid class = P) = \frac{(freq(I) + 1)}{(N_P + V)} = \frac{2}{13}$$

$$P(to \mid class = P) = \frac{(freq(to) + 1)}{(N_P + V)} = \frac{1}{13}$$

$$P(I \mid class = N) = \frac{(freq(I) + 1)}{(N_N + V)} = \frac{1}{12}$$

$$P(to \mid class = N) = \frac{(freq(to) + 1)}{(N_N + V)} = \frac{1}{12}$$

$$P(like \mid class = P) = \frac{(freq(like) + 1)}{(N_P + V)} = \frac{2}{13}$$

$$P(run \mid class = P) = \frac{(freq(run) + 1)}{(N_P + V)} = \frac{2}{13}$$

$$P(like \mid class = N) = \frac{(freq(like) + 1)}{(N_N + V)} = \frac{1}{12}$$

$$P(class = N \mid run) = \frac{(freq(run) + 1)}{(N_N + V)} = \frac{2}{12}$$

$$P(I\ like\ to\ run \mid class = P) = \prod p(t_i \mid class = P) = \frac{2}{13} \times \frac{2}{13} \times \frac{1}{13} \times \frac{2}{13} = 2.8 \times 10^{-4}$$

$$P(I\ like\ to\ run \mid class = N) = \prod p(t_i \mid class = N) = \frac{1}{12} \times \frac{1}{12} \times \frac{1}{12} \times \frac{2}{12} = 9.65 \times 10^{-5}$$

Como $P(I\ like\ to\ run \mid class = P) > P(I\ like\ to\ run \mid class = N)$, classificamos a frase "I like to run" como **P**.

II. Programming and critical analysis

6)

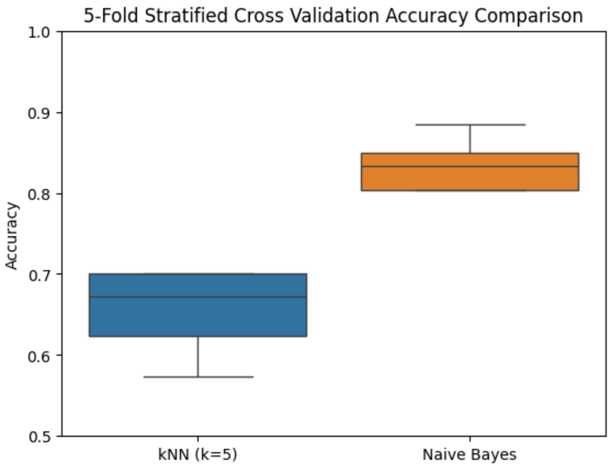
a) Ao analisar o boxplot abaixo, vemos que o modelo Naive Bayes apresentou uma precisão significativamente maior em comparação com o kNN, sendo a sua mediana das precisões próxima de 0.83 em comparação com 0.67 do kNN.

Adicionalmente, o modelo Naive Bayes aparenta ser mais estável já que tem uma menor dispersão dos dados no boxplot (caixa menor e intervalos interquartis mais pequenos), o que indica que a variação das precisões de cada fold é relativamente baixa. O kNN, por outro lado, apresenta uma variação maior nestas precisões, o que indica menos estabilidade.

Estes resultados podem dever-se ao facto de o kNN ser sensível à escala das variáveis, já que calcula distâncias para fazer previsões (ou seja, features como a idade, com intervalos de valores grandes, vão ter maior influência nos resultados e torná-lo menos preciso e estável), ao contrário do modelo Naive Bayes que, por ser um classificador probabilístico com suposições simples (independência condicional entre as features e distribuição Gaussiana para todas as variáveis), consegue manter uma performance mais estável mesmo com dados que não estejam perfeitamente normalizados ou escalados.

Mediana das precisões do modelo kNN: 0.6721

Mediana das precisões do modelo Naive Bayes: 0.8333



b) Comparando os resultados do modelo kNN antes e depois de fazer Min-Max Scaling, conseguimos perceber que este modelo beneficiou bastante deste escalonamento dos dados.

Primeiramente, a precisão melhorou bastante, já que a média das 5-folds está próxima de 0.82 e a mediana é agora 0.83(3), igual à do modelo Naive Bayes.

Para além disso, a variação da precisão também reduziu pois, os intervalos interquartis estão menores (até menores que do modelo Naive Bayes), significando que a estabilidade do modelo também aumentou considerávelmente. Isto faz todo o sentido uma vez que, se os dados estiverem escalados e normalizados, features com magnitudes muito elevadas já não vão ter um impacto tão grande no cálculo das distâncias necessário para classificar os dados com o modelo kNN.

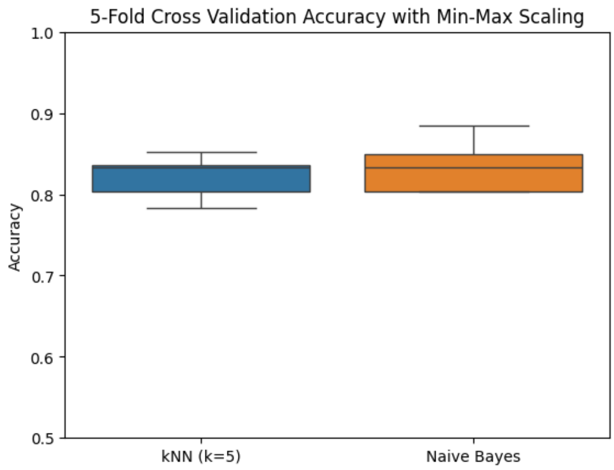
Quanto ao modelo Naive Bayes, vemos que o boxplot manteve-se inalterado, não mostrando nenhuma variação nas precisões dos 5-folds. Isto já era de esperar pois, como foi dito na questão anterior, o modelo Naive Bayes assume uma distribuição Gaussiana para as features e por isso, não é sensível ao escalonamento. Isto porque o Min-Max scaling apenas altera os intervalos de valores que as features podem tomar, mantendo as distribuição das mesmas.

Precisão média do modelo kNN: 0.8217

Precisão média do modelo Naive Bayes: 0.8350

Mediana das precisões do modelo kNN: 0.8333

Mediana das precisões do modelo Naive Bayes: 0.8333



c) Para testar a hipótese de que "o modelo kNN é estatísticamente superior ao Naive Bayes no que toca à precisão", realizamos um t_test, antes e depois de aplicar o min-max scaler, obtendo um p-value de 0.9987 e 0.7463, respetivamente.

Como em ambos os casos o p-value é maior que os valores habituais dos níveis de significâncias (neste caso 5%), rejeitamos a hipótese em questão, ou seja, a afirmação apresentada não é verdadeira.

Estatística t antes do Min-Max: -0.7271

p-value antes do Min-Max: 0.7463

Estatística t depois do Min-Max: -6.6903

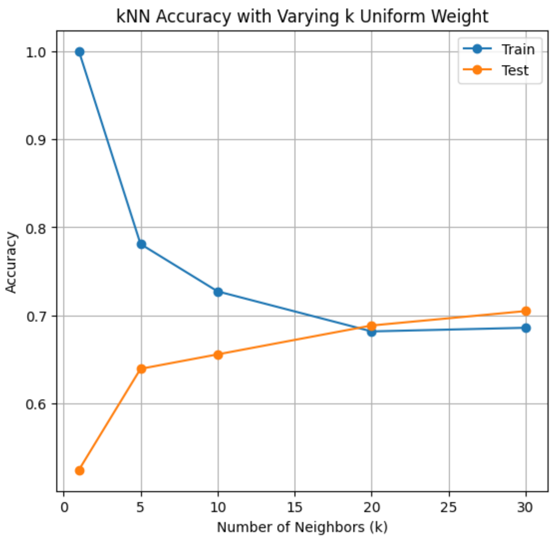
p-value depois do Min-Max: 0.9987

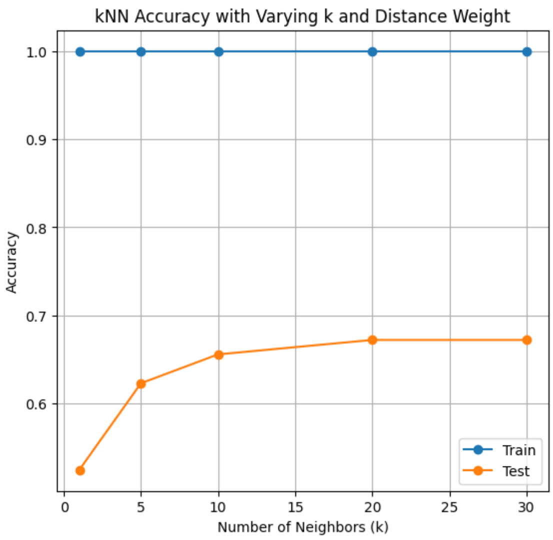
(antes do Min-Max) Rejeitamos a hipótese

(depois do Min-Max) Rejeitamos a hipótese

7)

a)





b) Aumentar o k num modelo kNN geralmente melhora a capacidade de generalização do mesmo. Isto pode ser explicado pois, para valores pequenos de k, o modelo ajusta-se demasiado aos dados de treino porque toma a decisão com base no(s) vizinho(s) mais próximo(s), mesmo que estes sejam outliers. Ou seja, temos overfitting porque este desempenho não se refletiria nos dados de teste. Por sua vez, para valores demasiado grandes de k, vai considerar vizinhos muito distantes e aos quais não tem relação. Torna-se um classificador muito simples e com underfitting. O ideal é encontrar um equilíbrio, de maneira que, o modelo considere um número de observações suficientemente grande para evitar capturar o ruído dos dados, mas não demasiado grande para não ter underfitting.

Quanto aos gráficos obtidos, comecemos por comentar a precisão do kNN para pesos uniformes (que classifica com base nos k vizinhos mais próximos sem dar prioridade a nenhum vizinho em particular, quer esteja mais perto ou mais longe):

- quando o k=1, a precisão do modelo para os dados de treino é 1 e para os dados de teste é perto de 0.5. Temos então um claro caso de overfitting e uma capacidade de generalização fraca devido ao que foi referido acima.

- à medida que o k aumenta (de 1 para 20), o modelo vai-se equilibrando e a sua capacidade de generalização aumenta (temos uma diminuição da precisão dos dados de teste e um aumento na precisão nos dados de treino).

- a partir de k=20, a precisão do modelo mantêm-se praticamente estável, com a precisão dos dados de treino a aumentar.

Podemos então ver que, para pesos uniformes, o aumento do k afeta o modelo como esperado, aumentando a generalização do mesmo (até um certo valor que, ainda que não tenha acontecido neste gráfico, espera-se que seja o caso).

No que toca à precisão do kNN, para pesos baseados na distância, temos um comportamento bastante diferente. Isto é facilmente observado no gráfico para os dados de treino que se mantém constante (com precisão = 1). A diferença é que, quando usamos pesos baseados nas distâncias, estamos a considerar os k vizinhos mais próximos, mas a dar mais importância aos mais próximos desses k. Ou seja, mesmo que aumentemos o k e, por isso, consideremos mais vizinhos, vamos dar sempre mais peso aos que estão mais próximos, o que faz com que o aumento do k não tenha tanto impacto neste modelo como no kNN com pesos uniformes. Outra maneira de ver isto é que, para os dados de teste, ainda que haja um aumento na precisão como no gráfico para pesos uniformes, este mantém-se constante a partir de k=20, significando que aumentar o k a partir daí não tem qualquer impacto na performance do modelo.

Podemos então concluir que o aumento do k, no geral, melhora a generalização do modelo kNN, sendo mais impactante no modelo que usa pesos uniformes do que no de pesos baseados na distância.

8) O modelo Naive Bayes apresenta duas dificuldades na aprendizagem do data set dado.

Primeiramente, o Naive Bayes assume que todas as variáveis são independentes umas das outras. Sendo o data set relacionado com “casos reais”, mais concretamente, doenças de coração e as possíveis causas, é muito plausível afirmar que estas causas estão muitas vezes correlacionadas, o que afeta bastante a precisão e desempenho do modelo.

Aliás, conseguimos confirmar este dado com o código apresentado abaixo que demonstra que de facto, existe correlação entre algumass features dadas.

Adicionalmente, o modelo Naive Bayes é melhor usado quando trabalhamos com features categóricas, o que no data set dado não é sempre o caso. Muitas variáveis como idade, nível de colesterol, ritmo cardíaco, entre outras, são variáveis contínuas. Para criar um modelo Naive Bayes que também aprenda data sets com variáveis contínuas, é comum supor que estas variáveis seguem uma distribuição Normal ou Gaussiana. Como é óbvio, esta suposição nem sempre é acertada, muito menos no contexto deste data set, o que pode fazer com que o modelo apresente dificuldades na aprendizagem do mesmo. Para além disso, este data set específico tem algumas variáveis categóricas que, por serem representadas por valores numéricos (por ex. por serem binárias), não são ignoradas nesta suposição da distribuição, o que também afeta os resultados.

Correlação entre age e trestbps: 0.2793509065612883

Correlação entre age e chol: 0.2136779565595618

Correlação entre exang e oldpeak: 0.28822280778276543

...

END