

Aprendizaxe automática

Exame de avaliación continua

Xoves, 8 abril 2021

Os números entre parénteses indican os puntos de cada pregunta (sobre unha nota total de 10)

1. (0.5) Que é a validación cruzada leave-one-out? Cando se debe empregar?

RESPOSTA: consiste en repetir N veces (sendo N o nº de patróns dispoñíbeis) o proceso de entrenamento e validación. En cada proba i, con $i=1..N$, entenas con tódolos patróns menos o patrón i, e testas sobre o patrón i, avaliando a calidade. Logo de repetir as N probas, promedias as N medidas de calidade para obter a calidade final. Débese empregar con N baixo, porque neste caso non hai patróns suficientes para facer unha validación K-fold cos K típicos ($K=4,5,10$).

2. (1) Dados os clasificadores e problemas da seguinte táboa co valor de Kappa (en %) en problemas de clasificación, calcula o seu ranking de Friedman.

Kappa(%)	P1	P2	P3
C1	23.5	33.1	65.2
C2	80.2	22.0	78.4
C3	15.8	18.4	56.2
C4	67.1	25.9	43.1
C5	55.2	21.7	60.2

RESPOSTA: táboa de rankings:

Posición	P1	P2	P3
1	C2	C1	C2
2	C4	C4	C1
3	C5	C2	C5
4	C1	C5	C3
5	C3	C3	C4

Táboa de ranking promedio:

	P1	P2	P3	Promedio
C1	4	1	2	2.3
C2	1	3	1	1.7
C3	5	5	3	4.3
C4	2	2	5	3.0
C5	3	4	3	3.3

Ranking de Friedman:

Posición	Clasificador	Ranking
1	C2	1.7
2	C1	2.3
3	C4	3.0
4	C5	3.3
5	C3	4.3

3. (0.5) Tes un clasificador con 2 hiper-parámetros sintonizábeis e queres facer unha validación cruzada 5-fold. Indica cantas particións necesitas e para que é cada unha.

RESPOSTA: con validación cruzada 5-fold necesitas 5 grupos de particións; se tes que sintonizar hiper-parámetros, debes usar tres particións (entrenamento, validación e teste) para cada fold. Polo tanto, necesitas $5 \times 3 = 15$ particións. Na sintonización de parámetros usas as particións de entrenamento e validación de cada fold para entrenar e testear, respectivamente. No teste, usas as particións de entrenamento e validación para entrenar, e a de teste para testear, en cada fold.

4. (1) Para que serve un teste de Wilcoxon? Como funciona?

RESPOSTA: serve para comparar dous algoritmos sobre un conxunto de problemas. Para isto comparas os dous vectores coas medidas de calidade dos dous métodos sobre o conxunto de problemas. O teste determina se a diferenza entre eles é estadísticamente significativa (dito técnicamente, se as distribucións das que proveñen os dous vectores teñen a mesma media). A hipótese nula é que si teñen a mesma media. O teste devolve un p-value, que se é menor que 0.05 fai que a hipótese nula sexa rexeitada (é dicir, que non teñen a mesma media, polo tanto son distintas, polo tanto a diferenza entre eles é estatísticamente significativa).

5. (0.5) Nun problema de clasificación multiclase, o LDA proxecta por defecto a un espazo de que dimensión? A que dimensións podería proxectar?

RESPOSTA: nun problema de C clases, proxecta por defecto a un espazo $(C-1)$ -dimensional. Pode proxectar a espazos de dimensión d con $d=1,2,\dots,C-1$.

6. (1) Que mide a matriz de covarianza dunha clase?

RESPOSTA: a varianza das entradas dos patróns desta clase a respecto da media de clase.

7. (1) Como se calcula a matriz de covarianza inter-clase nun problema de clasificación binaria?

RESPOSTA: como o produto exterior da diferenza entre as medias de clase por si mesma.

8. (1) Cando se presenta un patrón de entrada $\mathbf{x}$ a un nodo dunha árbore de decisión con 5 nodos fillo, que fai este nodo?

RESPOSTA: o nodo colle a súa entrada i asociada e compara x_i cos 4 umbrales asociados ao nodo, decidindo en cal dos 5 intervalos (tantos como nodos fillos) cae x_i , e pasando x_i ao nodo seleccionado.

9. (0.5) Como actúan as árbores de decisión sobre entradas continuas?

RESPOSTA: discretiza a entrada continua usando os umbrais que ten asociados esa entrada, e logo opera sobre ela como con calquera entrada discreta.

10. (0.5) Que mide a entropía dun nodo nunha árbore de clasificación?

RESPOSTA: mide en que grao os patróns que acadan ese nodo pertencen a clases distintas. A entropía vale 0 cando tódolos patróns dese nodo son da mesma clase (neste caso, é un nodo terminal), e aumenta de valor canto máis repartidos estén os patróns entre as distintas clases.

11. (1) Se temos unha única neurona con función de activación escalón, en que rexión do espazo de entrada debe estar $\mathbf{x}$ para que a saída z do perceptrón sexa +1? Xustifica a resposta.

RESPOSTA: a saída z vale +1 cando o patrón $\mathbf{x}$ ten unha proxección positiva na dirección do vector director $\mathbf{w}$ do perceptrón.

12. (0.5) Que forma xeométrica teñen as fronteiras de separación entre clases aprendidas por unha rede neuronal perceptrón multicapa? Por que?

RESPOSTA: son fronteiras lineares a cachos, xa que cada neurona oculta realiza unha separación linear do seu espazo de entrada, e a combinación de tódalas neuronas da lugar a unha fronteira complexa pero composta por cachos lineares (hiperplanos en n dimensións).

13. (1) En que consiste o algoritmo de entrenamento RMSProp para o MLP?

RESPOSTA: en escalar o momento pola súa covarianza, calculada a partires do cadrado do incremento no peso $\mathbf{w}$ ou no sesgo b .