

Clasificación I

Test de Lectura

Como discutiesemos, estos *tests de lectura* son una serie de preguntas simples que permiten monitorear si están siguiendo efectivamente las clases. No serán corregidos ni evaluados, los administrarán directamente ustedes. Cualquier duda que tengan será discutida en nuestras sesiones sincrónicas, la semana siguiente a publicado el material, luego de que hayan hecho un esfuerzo en trabajarlo, no antes. Es muy importante que trabajen este material y lo discutan entre ustedes.

1. Escribi la regla de Bayes para el caso de solo dos categorías mutuamente excluyentes. Repasa la regla de Bayes de cuando la hayas estudiado.
2. Supongamos un modelo de ‘credit scoring’ en donde el evento a predecir es si una persona pagara un crédito. Si los datos disponibles son un indicador binario de si el crédito se pago, y el único predictor X es la edad (en años). Intuitivamente, ¿qué es $f(X = 35|Y = 1)$?
3. En el ejemplo anterior, ¿qué es κ ?
4. V o F: en el ejemplo anterior el estimador de κ es la proporción de gente que pago el crédito.
5. V o F: como en logit o un modelo de regresión, en el modelo de análisis discriminante X puede ser cualquier variable, continua o discreta.
6. Si una de los predictores es el género de la persona, ¿cuál es el problema con análisis discriminante en este caso?
7. V o F: predecir que $Y = 1$ es equivalente a verificar que el logaritmo del ‘odds/ratio’ de esa observación es mayor que cero.
8. Si para el caso simple discutido anteriormente, el modelo logit es $p = F(\beta_1 + \beta_2 X)$. Supongamos que $\beta_1 = -50$ y $\beta_2 = 2$, ¿cuál es la edad a partir de la cual a la persona se le otorga el crédito?
9. Verifica todos los pasos para obtener la función de discriminación para el caso de varianzas iguales. ¿Por qué la función es lineal?
10. Obtene la función de discriminación para el caso de varianzas no iguales. Muestra que la función de discriminación es ahora cuadrática.
11. Supone que lo que queremos predecir es si una persona mirara una serie, una película o un documental, y que el predictor es el ingreso.

Supone que lo que quieres predecir es que genero de pelicula mirara una persona que gana 100,000 por mes. Explica como funciona el predictor en base a vecinos cercanos, con $K = 20$.

12. Por que aumentar el umbral c aumenta los verdaderos positivos pero tambien aumenta los falsos positivos?
13. V o F: cuando $c = 1$ no hay falsos positivos.
14. V o F: el punto del origen de la curva ROC se corresponde con $c = 0$
15. Por que ROC tiene pendiente positiva?
16. Pensa el argumento que esta en la nota al pie del slide que dice 'La curva ROC, formalmente'
17. Demuestra el resultado de la igualdad del slide que se titula 'El Clasificador Invertido'
18. V o F: debajo de la recta de 45° $tvp > tfp$.
19. Dibuja la curva ROC ideal.
20. V o F: si para los mismos datos de entrenamiento logit tiene su curva ROC siempre por debajo de vecinos cercanos, entonces no tiene sentido usar logit.
21. Para pensar: si para dos clasificadores (logit y vecinos cercanos, por ejemplo), sus curvas ROC se cruzan, que clasificador te parece que hay que usar?