

Instrucciones (130 puntos) Encierra en un círculo la respuesta o respuestas correctas. No hay preguntas capciosas.

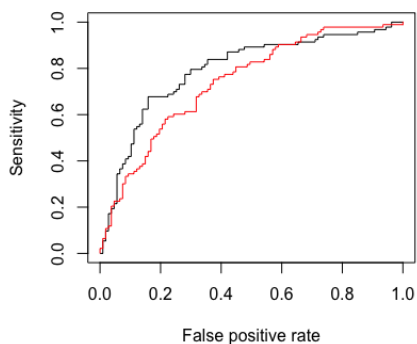
- (5^{pts}) 1. En un modelo de regresión logística con dos entradas, ajustamos por descenso en gradiente los coeficientes $\beta_0 = 0.5, \beta_1 = 1$ y $\beta_2 = -1$. ¿Cuál es la probabilidad estimada de positivo para las entradas $x_1 = 2, x_2 = 1$?
- (a) 0.50 (b) 0.62 (c) 0.82 (d) 0.95
- (5^{pts}) 2. En un modelo de regresión lineal, consideramos un modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, con coeficientes $\beta_0 = 10, \beta_1 = -1$ y $\beta_2 = 2$. ¿Cuál es el error cuadrático medio de este modelo sobre la siguiente muestra de prueba?

```
data.frame(x_1=c(0,-1,0), x_2=c(1,2,0), y=c(7,2,10))
```

```
##   x_1 x_2 y
## 1    0  1  7
## 2   -1  2  2
## 3    0  0 10
```

- (a) 64.67 (b) 189.57 (c) 0 (d) -35.21
- (8^{pts}) 3. Supongamos que tenemos un problema de mínimos cuadrados que queremos resolver usando descenso en gradiente. ¿Cuál de las siguientes son verdaderas?
- (a) Si en las primeras iteraciones de descenso en gradiente la función objetivo (RSS) crece, puede ser que el tamaño de paso es demasiado grande.
- (b) En descenso en gradiente cada iteración reduce el valor de la función objetivo.
- (c) Si no se usa regularización, el algoritmo puede converger a soluciones con distintos valores mínimos de RSS (la función objetivo).
- (d) Si el tamaño de paso es muy chico, entonces puede ser que descenso en gradiente tome muchísimas iteraciones para alcanzar el mínimo.
- (8^{pts}) 4. ¿Por qué escalamos las entradas antes de usar descenso en gradiente para regresión lineal o regresión logística? Selecciona todas las que apliquen.
- (a) Es necesario porque si no escalamos el algoritmo se puede atorar en un mínimo local con valor relativamente alto de la función objetivo.
- (b) Porque hace más rápido el cálculo de cada iteración de descenso en gradiente.
- (c) Porque así se requieran menos iteraciones para alcanzar una buena solución.
- (d) Porque así se obtiene una solución que ajusta mejor a los datos (menor RSS).
- (5^{pts}) 5. Supongamos que tenemos el modelo $f(x) = \beta_0 + \beta_1 x_1$ (regresión lineal, respuesta numérica), con dos datos de entrenamiento $(0, 0), (1, 1)$. Si en la primera iteración de descenso en gradiente ponemos $\beta_0 = 0, \beta_1 = 0$ y tomamos como tamaño de paso $\eta = \frac{1}{2}$, ¿Cuál es el valor (más cercano) de la siguiente iteración para β_1 ? (usa RSS dividido por el número de casos de entrenamiento como función objetivo).
- (a) 1 (b) -0.5 (c) 0.5 (d) 0.25 (e) -0.25

- (5^{pts}) 6. El máximo de ozono que se observa en la Ciudad de México un día dado se mide en partes por millón. Supón que quieres construir un modelo predictivo del máximo de ozono para el día de mañana. Este es un problema de
- (a) Regresión (b) Clasificación
- (6^{pts}) 7. Considera las curvas ROC de prueba de los siguiente dos modelos de clasificación binaria:



¿Cuál de las siguientes es verdadera?

- (a) Si tuviéramos que escoger un solo modelo (sin más información), preferiríamos el negro sobre el rojo.
- (b) Si queremos sensibilidad de 60 %, con el modelo negro podemos obtener una especificidad entre 0 y 20 % para la muestra de entrenamiento.
- (c) El modelo negro tiene menor error de entrenamiento que el modelo rojo.
- (8^{pts}) 8. En un modelo de regresión logística con regularización, escoge todas las que aplican:
- (a) Disminuir al regularización resulta en mejor desempeño en la muestra de prueba.
- (b) Agregar una nueva entrada al modelo siempre resulta en mejor o igual desempeño en la muestra de prueba.
- (c) Agregar una nueva entrada al modelo siempre resulta en mejor o igual desempeño en la muestra de entrenamiento.
- (d) Disminuir la regularización no empeora el desempeño en la muestra de entrenamiento.
- (5^{pts}) 9. ¿Cuál es la predicción de 2 vecinos más cercanos en $x = 0$ para los siguientes datos?

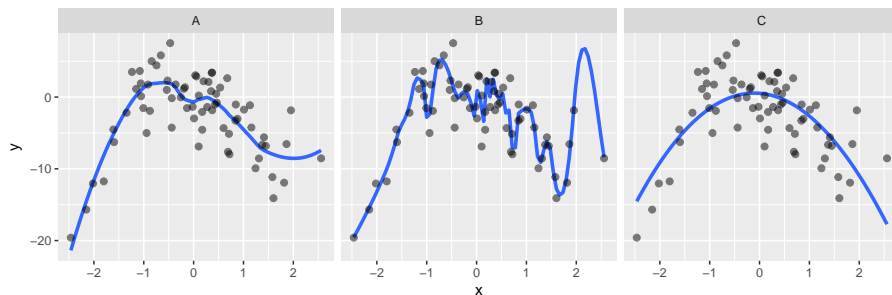
```
##      x y
## 1  1.47 2
## 2 -0.65 4
## 3  0.50 3
## 4  0.06 4
## 5 -1.24 3
```

- (a) 2.5 (b) 3 (c) 3.5 (d) 4
- (8^{pts}) 10. Tenemos 5000 variables de entrada con 50 casos de entrenamiento para predecir una variable binaria. Primero seleccionamos las 30 variables con mayor correlación con la respuesta, y hacemos regresión logística con estas 30 variables. Estimamos el error de clasificación por validación cruzada-10 para estas 30 variables y el modelo obtenido. Más tarde nos dicen que la variable respuesta estaba mal, y que fue permutada al azar (con probabilidad de positivo igual a 0.5).
- (a) Si usamos una muestra de prueba suficientemente grande, el error estimado de prueba de este modelo debería ser cercano a 0.5 (al azar).
- (b) La estimación del error por validación cruzada según el proceso de arriba debería estar cercana 0.5.

- (c) El proceso de selección de las 30 variables debería estar incluido en la validación cruzada.
- (d) La estimación del error por validación cruzada descrita arriba será demasiado optimista porque en la selección de variables usamos todos los datos de entrenamiento.
- (5^{pts}) 11. Ajustaste dos modelos de regresión lineal regularizada con ridge sin ordenada al origen ($\beta_0=0$) con distintos valores de regularización $\lambda = 0$ y $\lambda = 1$. Los coeficientes de estos dos modelos son $\beta^1 = (22.1, 5.6)$ y $\beta^2 = (7.8, 1.2)$, pero no te dicen cuál corresponde a cuál parámetro λ .
- (a) β^1 corresponde a $\lambda = 1$ y β^2 corresponde a $\lambda = 0$
- (b) β^1 corresponde a $\lambda = 0$ y β^2 corresponde a $\lambda = 1$
- (c) No se puede saber.
- (8^{pts}) 12. ¿Cuáles de las siguientes son verdaderas para regresión lasso?
- (a) Cuando regularizamos con λ más grande, el error de entrenamiento no decrece y el error de prueba puede bajar.
- (b) Cuando regularizamos con λ más grande, incrementamos potencialmente el número de variables que entran en el modelo.
- (c) Podemos escoger la regularización λ óptima de forma que el error de entrenamiento (sin penalizar) sea lo más chico posible.
- (d) Podemos escoger la regularización λ óptima de forma que el error de validación (sin penalizar) sea lo más chico posible.
- (5^{pts}) 13. Supón que en un problema de clasificación binaria tenemos un modelo que predice, para una muestra de prueba de tamaño $n = 3$, probabilidades de positivo 0.4, 0.2, 0.8. Observamos que las clases son *neg, pos, pos*. ¿Cuál de las siguientes está más cerca de la devianza de prueba de este modelo (divide entre el número de casos y usa el factor de 2)?
- (a) 1.56 (b) 1.83 (c) 0.91 (d) 2.71
- (8^{pts}) 14. Supongamos que para un problema de clasificación binaria, construimos una estimación de la probabilidad de clase 1 $\hat{p}_1(x)$. Consideramos entonces dos clasificadores G_1 y G_2 , contruidos a partir de $\hat{p}_1(x)$ con puntos de corte 0.2 y 0.5 respectivamente. Sobre una muestra de prueba, ¿cuáles de los siguientes son verdaderos?
- (a) La tasa de falsos positivos de G_2 es mayor o igual que la de G_1 .
- (b) La tasa de falsos negativos de G_2 es mayor o igual que la de G_1 .
- (c) La sensibilidad de G_2 es mejor o igual que la de G_1 .
- (d) La especificidad de G_2 es mejor o igual que la de G_1 .
- (10^{pts}) 15. Supón que estamos construyendo un modelo predictivo (puedes pensar en regresión logística). El error de entrenamiento es muy similar al error de prueba, y la muestra de prueba es muy grande. ¿Cuáles de las siguientes pueden mejorar el desempeño predictivo?
- (a) Incrementar la regularización.
- (b) Agregar variables derivadas de las que están en el modelo actualmente.
- (c) Eliminar algunas variables que ya están en el modelo.
- (d) Reducir la regularización.
- (e) Usar una muestra de prueba más chica.
- (10^{pts}) 16. Supón que estamos construyendo un modelo predictivo (puedes pensar en regresión logística). El error de entrenamiento es casi cero, pero el error de prueba (con una muestra de prueba grande) es considerablemente alto. ¿Cuáles de las siguientes pueden mejorar el desempeño predictivo?
- (a) Incrementar la regularización.
- (b) Agregar variables derivadas de las que están en el modelo actualmente.

- (c) Eliminar algunas variables que ya están en el modelo.
- (d) Reducir la regularización.
- (e) Evaluar con los datos de prueba donde las predicciones tienen error más bajo.

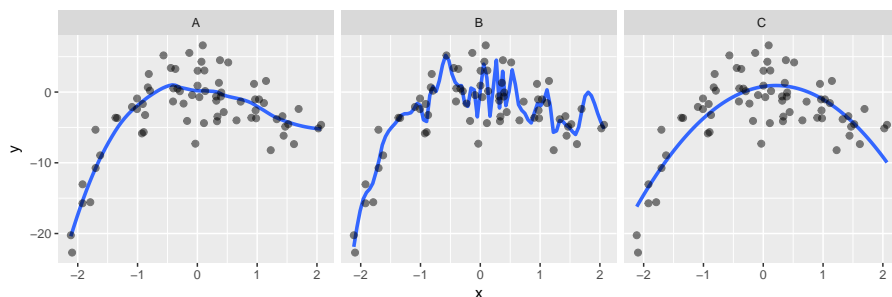
(5^{pts}) **17.** Considera los siguientes ajustes:



¿Cuáles de ellos tienen mayor error de predicción?

- (a) A y B
- (b) A y C
- (c) B y C

(5^{pts}) **18.** Considera los siguientes ajustes:



¿Cuáles de ellos tienen mayor error de entrenamiento?

- (a) A y B
- (b) A y C
- (c) B y C

(6^{pts}) **19.** ¿Cuáles de las siguientes son verdaderas?

- (a) Hacer validación cruzada con 2 vueltas generalmente es mala idea porque los modelos que probamos pueden ser muy diferentes al que ajustamos con todos los datos.
- (b) Si tenemos 500 datos de entrenamiento, hacer validación cruzada con 500 vueltas generalmente es mala idea porque los modelos que probamos pueden ser muy diferentes al que ajustamos con todos los datos.
- (c) Si tenemos 500 datos de entrenamiento, hacer validación cruzada con 500 vueltas puede ser demasiado costoso computacionalmente.

(5^{pts}) **20.** Un clasificador tiene especificidad de 0.9 y sensibilidad de 0.5 sobre una muestra de prueba. Si la muestra de prueba contiene 0.25 casos positivos, la tasa de clasificación correcta de este clasificador es de

- (a) 0.5
- (b) 0.6
- (c) 0.7
- (d) 0.8
- (e) Ninguna de las anteriores