

Regularizacion. Lasso y Ridge

Test de Lectura

Como discutiesemos, estos *tests de lectura* son una serie de preguntas simples que permiten monitorear si estan siguiendo efectivamente las clases. No seran corregidos ni evaluados, los administraran directamente ustedes. Cualquier duda que tengan sera discutida en nuestras sesiones sincronicas, 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. Por que un modelo con 1 predictor no es un caso particular de uno con dos predictores?
2. Cual es el problema con buscar exhaustivamente en todos los modelos posibles? En la EPH hay unos 50 predictores de la pobreza. Cuantos modelos habria que estimar si la busqueda fuese exhaustiva?
3. Por que forward selection no es un metodo exhaustivo?
4. Si hay 1.000 observaciones y 1.200 predictores, cual es la ventaja de *forward* selection? (y cual es el problema con *backward* selection?).
5. Por que decimos que ‘tachar variables es una forma extrema de achicar sus coeficientes en el modelo’?
6. Por que no se puede minimizar la funcion objetivo de LASSO usando los metodos habituales de derivar e igualar a cero?
7. Reproduci todos los detalles del grafico del slide 13.
8. Verifica todas las cuentas del slide 14.
9. V o F: en el caso simple estudiado, si la pendiente de la funcion de penalidad es positiva (hacia la derecha), la solucion de LASSO es $\beta_l = 0$.
10. Cuanto vale λ^* ?
11. V o F: para el caso simple, λ^* es el doble del coeficiente estimado por MCO.
12. Por que en LASSO, cuando la solucion es interior decimos que esta ‘corrida hacia cero’ en relacion al metodo de MCO.

13. Intuitivamente, por que los resultados de LASSO son sesgados, en relacion a MCO?
14. Repasa el argumento de porque si en el margen el beneficio de mover un coeficiente de cero es mas chico que el costo de hacerlo, entonces no conviene introducir esa variable explicativa.
15. Porque en LASSO el costo marginal de introducir variables es constante mientras que el beneficio de hacerlo es positivo pero decreciente?
16. Revisa los detalles algebraicos del caso de ridge.
17. Pensa el argumento de ‘costo/ingreso marginal’ pero para *ridge* y explica por que ahora siempre la solucion es interior (tip: ahora ambos son variables).
18. V o F: en *ridge* siempre existe un valor de λ para el cual su error cuadratico medio esta por debajo del de MCO.
19. Intuitivamente, explica porque LASSO/Ridge son estrategias sesgadas que potencialmente le ‘ganan’ a MCO en capacidad predictiva.
20. Explica como elegirias λ por cross validation.