Examen Parcial: Aprendizaje Estadístico

EST-25134, Primavera 2021
Dr. Alfredo Garbuno Iñigo

1. Instrucciones

- Sus respuestas deben ser claras y deben poder explicar los resultados. Incluyan también sus procedimientos de manera ordenada. En la sección práctica incluye comentarios en el código.
- Se resolverán dudas en sesiones de 45 minutos después de cada clase en los días 23 y 25 de Marzo. Es decir, de 5:30pm a 6:15pm.
- No pueden compartir soluciones entre diferentes equipos.
- Al entregar este examen afirmas que el trabajo se realizó sólo con tus compañeros de equipo. El material que utilizaste para apoyarte consistió de las notas en clase (pdfs del pizarron virtual que se comparten en canvas), y el código fuente se compartió en el repo de Github.

  Al entregar estás dando tu consentimiento para que bajo sospecha y suficiente evidencia de copia se anule su evaluación.
- Por favor, incluyan las citas relevantes en caso de utilizar material adicional para el desarrollo de sus respuestas.

1.1. Entrega

Enviar por correo electrónico la carpeta comprimida `equipo-xx.zip` con el asunto **`[AE-2021] Examen Parcial - XX.`** Donde XX representa el identificador de tu equipo. El cuerpo del mensaje debe incluir los nombres y claves únicas de los integrantes de los equipos. La resolución del ejercicio teórico, datos y codigo de solución deberá estar incluido en paquete de entrega. La fecha límite para entregarlo es el viernes 26 de Marzo a las 11:59pm. No se aceptarán entregas extemporáneas. Será mejor entregar un examen resuelto parcialmente, que no entregar nada.

1.2. Ponderación

La pondereción es la siguiente: 30 pts (teórico), 70 pts (práctico) y 5 pts (si la entrega de la sección teórica la realizan en LaTeX).

2. Teoría

1. En clase establecimos el resultado siguiente para aprendizaje CLA (Convexo, Lipschitz y Acotado):

   Teorema 2.1. *Supongamos que la función de pérdida es convexa y ρ-Lipschitz. Entonces aplicar el principio ERM con un regularizador $\lambda\|w\|_2^2$ es estable (es decir, estable en promedio con reemplazos individuales) con tasa $\frac{2\rho^2}{\lambda m}$. De donde concluimos que*

   $$\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S)) - L_S(A(S))] \leq \frac{2\rho^2}{\lambda m}\,.$$

   En este apartado probarán que podemos llegar a un criterio semejante en el caso CSA (Convexo, Suave y Acotado). Para esto revisa la prueba que hicimos en clase para el teorema anterior. Ahora, prueben el teorema siguiente.

   Teorema 2.2. *Supongamos que la función de pérdida es convexa, β-suave y no-negativa. Entonces aplicar el principio ERM con un regularizador $\lambda\|w\|_2^2$ es estable (es decir, estable en promedio con reemplazos individuales) con tasa $\frac{48\beta K}{\lambda m}$. Donde $K = \mathbb{E}[L_S(A(S))]$. De donde concluimos que*

   $$\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S)) - L_S(A(S))] \leq \frac{48\beta K}{\lambda m}\,.$$

   Pista: utiliza la propiedad de auto-acotamiento de una función β-suave y no-negativa en conjunción con el Corolario de la Sección 1 en las notas de la Clase-14. Es decir, utiliza

   $$\|\nabla f(w)\|^2 \leq 2\beta f(w)\,,$$

   y la cota superior

   $$f(u) \leq f(w) + \nabla f(w)^\top (u-w) + \frac{\beta}{2}\|u-w\|^2\,.$$

2. El resultado anterior puede refinarse un poco si además asumimos que la función de pérdida está acotada como función de una observación. Es decir, si asumimos que $\ell(0,z) \leq C$ con $C > 0$ para toda $z \sim \mathcal{D}$.

   a) ¿Cómo cambian las conclusiones que probaron anteriormente bajo este supuesto adicional?

   b) ¿Cómo se relaciona este resultado con aprendizaje no-uniforme?

3. Práctica

Los ejercicios de la sección práctica los podrás encontrar en el sitio de Canvas de la materia en el anuncio correspondiente. El zip incluye instrucciones y datos. Tu carpeta de entrega deberá contener tu código en el archivo `.Rmd` y un archivo `.csv` con las predicciones de tu modelo.