

Reporte | Carlos Castro Correa 103531

Para mi segundo reporte no teníamos opciones para escoger así que mi reporte trata sobre: "Convex Optimization for Big Data".

En lo personal, encontré mucho sentido en este reporte porque creo que lo que comenta tiene que ver mucho con los retos que enfrentamos como matemáticos profesionales, en mi caso, al menos en el ITAM aprendimos una serie de algoritmos y opciones para resolver problemas reales (como el planteamiento de problemas de optimización convexa), sin embargo, la mayoría de los problemas escolares y académicos son con menos datos o baja escala, si bien el enfoque teórico es adecuado, cuando tratas de resolver un problema real te encuentras con dificultades como: datos faltantes, los datos no caben en la memoria, nadie te compra MATLAB, etc. En otras palabras, salen con un déficit que tienes que cubrir por tu cuenta porque los problemas reales tienen condiciones menos estables y debes encontrar una solución.

Respecto al paper, hace mucho énfasis en la adaptación de técnicas actuales de optimización convexa tradicionales para ajustarlas a retos de muchos datos. Podemos resumir estas adaptaciones en tres rubros:

- Condiciones de primer orden.
- Aleatorización.
- Cómputo paralelo y distribuido.

Condiciones de primer orden.

En mi opinión, esta opción es la que menos se aleja del planteamiento original del problema pues creo que en lugar de buscar nuevas tecnologías o herramientas estadísticas para solucionar el problema, plantea modificaciones a la función original para llegar al óptimo,

Con la finalidad de no solo hacer un resumen de lo que dice el *paper*, me gustaría mencionar que encuentro en esta sección algo parecido a lo que ocurre en el problema de regresión lineal. Al usar esta técnica podemos encontrar algunas dificultades debido a la escala de los datos y otras condiciones que aumentan la varianza de nuestros estimadores; en este caso, la regresión Lasso y Ridge se encargan de ajustar el problema agregando un parámetro λ a la función objetivo para hacer que la varianza de nuestra estimación sea más estable. Como lo mencionaba no se trató de buscar otro programa o tecnología para resolver el problema, más bien basarse en la idea general, modificando algunos detalles que nos ayudarán para encontrar la solución.

Como matemáticos, me parece que estamos muy acostumbrados a este enfoque más que a recurrir a nuevas tecnologías, siempre pensamos en planteamientos alternativos o modificaciones que sean útiles.

Aleatorización.

Además de las modificaciones que podemos realizar a las condiciones de primer orden y que nos sirvan para atacar a los problemas de gran escala, en ocasiones no son suficientes porque estas soluciones requieren que los algoritmos iteren de muchas veces y nos enfrentamos a problemas más graves. En este caso podemos utilizar aleatorización.

Esta sección me recuerda mucho al tema de heurísticas: una serie de pasos que si bien, no garantizan que podemos encontrar el óptimo exacto, si podemos ir acercándonos poco a poco a la solución en función de los parámetros iniciales que consideremos.

En el documento se menciona un enfoque muy conocido, la adaptación del gradiente de forma estocástica. En ocasiones, cuando aplicamos descenso en gradiente la convergencia a la solución es muy lenta o de

repente se estanca en algún valor, en este caso, a veces es bueno buscar un poco en otra dirección u otro tamaño de paso de vez en cuando, si obtenemos una mejora seguimos en la dirección y si no, buscamos por otro lado. Como lo mencionaba, esta idea podría ser considerada una heurística, en particular las heurísticas que aplicamos en problemas de investigación de operaciones o en sistemas complejos adaptativos.

Cómputo paralelo y distribuido.

Este me parece el método más ingenieril y moderno que podemos encontrar para solucionar este tipo de problemas.

Este enfoque plantea escalar la capacidad computacional que tenemos para procesar nuestros datos en lugar de modificar o adaptar nuestros criterios analíticos. Se trata de agregar capacidad de procesamiento mediante un clúster de máquinas y un nuevo paradigma de repartición de tareas basado en la noción de nodos maestros y esclavos.

Esta idea podría parecer muy novedosa y aunque en gran parte lo es, lo primero en lo que pienso cuando escucho cómputo paralelo y distribuido es en el planteamiento de problemas de **programación dinámica**: dividir el problema en etapas de tal forma que podemos encontrar la solución de cada una y al final podemos encontrar la solución general utilizando el óptimo de cada etapa o más fundamental, en algoritmos como QR y otros de álgebra lineal que nos muestran las ventajas de dividir las tareas en pequeños problemas con soluciones parciales.

A pesar de que los primeros dos enfoques de solución nos ayudan en la mayoría de los casos, el cómputo en paralelo y servicios como AWS, Digital Ocean, entre otros, se han convertido en herramientas fundamentales para científicos de datos, sobre todo considerando que modificar las condiciones de primer orden o aleatorización requieren pensar el planteamiento de una nueva idea y eso lleva tiempo, en la vida real generalmente necesitas una solución rápida y efectiva por lo que, en ocasiones, tenemos que optar por el recurso del cómputo en paralelo sin tener mucho tiempo para replantearlo.