Santiago Ahumada Lozano.

Guión video Proyecto final.

1. (The Optimization Problem and some constraints of the gradient descent):

¡Hola! Mi nombre es Santiago Ahumada Lozano. Hoy les voy a hablar sobre el problema de optimización y algunas limitaciones del método del **Gradiente en descenso**. Esta es la entrega del proyecto final de la asignatura **Introducción al Aprendizaje de máquina**, dictada por el profesor Francisco Gómez.

1. (What is the optimization about?)

Como bien hemos aprendido en cursos básicos de cálculo, en la industria resulta de gran atractivo encontrar valores fijos de ciertos recursos con el objetivo de minimizar el coste y maximizar la ganancia, por ejemplo.

En este caso podremos definir a un problema de optimización de la siguiente manera: Sea f una función de cierto subconjunto D, contenido en un hiperplano d-dimensional en los reales. Entonces la solución del problema de optimización esta dada por encontrar bien el valor mínimo alcanzado por dicha función o bien el argumento que minimiza esa función. (Breve explicación en tablero).

1. (Optimization and machine learning ):

A simple vista, parece que la optimización y el aprendizaje máquina son ramas de la matemática y de la computación bien lejanos. Sin embargo, según se sabe, todo problema de aprendizaje trae implícito uno de optimización que lo haga eficiente en términos de complejidad.

Por ejemplo en redes Neuronales, resulta necesario, después de tener planteada cierta red profunda de neuronas que genere bajos errores dentro y fuera de la muestra. Entonces entra la función coste a ser minimizada, hallando los bias y los pesos adecuados que minimizan el error en el aprendizaje. Aquí algunos casos:

1. (Famous Problems):

Entre problemas teóricos y aplicados, sin lugar a dudas resaltan dos problemas NP-duros, uno de ellos, tratado en el curso, es el problema de clasificación centrado en minimizar la media sobre los errores de una hipótesis sobre un conjunto muestra de datos (Conocido como **in-sample error**). Otro, es conocido como el **problema del viajante** o TSP (Por Traveller salesman problema). Se trata de hallar el camino que pasa una vez por una cantidad fija de nodos (ciudades) que minimice la distancia total recorrida.

1. (Popular archievements)

En la práctica, resolver problemas NP-duros y reducirlos a complejidad, por ejemplo, polinomial, implican resolver un problema que concede el éxito empresarial a determinados emprendimientos, este es el caso de Uber y su problema de minimizar el camino que conecta dos nodos en cierto grafo.

Adicional a ello, la optimización de recursos ha sido mandatoria en la ingeniería civil moderna a la hora de minimizar la cantidad de materiales, esta implementación ha sido sin duda utilizada en la construcción del nuevo estadio del Real Madrid, el Santiago Bernabéu. El cual será estrenado este mes de septiembre.

1. (Recent advances)

A diferencia de teorías netamente puras, como la teoría de la generalización, cuyos avances se dan en términos de que tan próximos estamos de encontrar mejores y mejores explicaciones de que tan capaces deberían ser las máquinas de aprender fuera de la muestra (O generalizar) en términos del tamaño del conjunto de entrenamiento y de cierta tolerancia de error, la optimización encuentra avances a medida que resuelve problemas NP-duros, encuentra nuevos algoritmos y desarrolla teoría que genere nuevas herramientas y perspectivas para abordar cierto conjunto de problemas.

1. (Gradient Descent)

El método de optimización por excelencia del aprendizaje máquina (Breve explicación en tablero)

1. (Constraints)

(Explicación con código)

En resumen, las limitaciones principales del gradiente en descenso son

* La lentitud en la convergencia
* Alta sensibilidad al coeficiente de tasa de aprendizaje
* El algoritmo se acompleja cuando el conjunto de datos es muy grande o sus datos presentan múltiples valores
* Encuentra el mínimo local mas cercano en lugar del mínimo global

Es por ello que se han desarrollado variaciones de este algoritmo

1. (1. Momentum)

El primero de ellos, se denomina momento y esta dado por esta regla de actualización. El coeficiente beta se denomina momento y optimiza la manera en que combinamos linealmente el vector anterior y su gradiente en descenso.

1. (2.Nasterov Momentum Update)

O también conocido como gradiente en descenso con momento acelerado. Es una implementación propuesta por el matemático (Aún vivo) Yurii Nasterov, el cual considera el reescalamiento de los puntos anteriores también al evaluar el gradiente.

1. (AdaGrad)

En este caso, la variante a considerar, es que el coeficiente de la tasa de aprendizaje será dependiente del paso. Y estará definido por esta formula (Breve explicación de los términos).

Sin lugar a dudas existen más algoritmos como AdaDelta, Adam, Gradiente estocástico, e incluso algoritmos ya vistos en el curso como el anti-sobreajuste regularización o las máquinas de soporte vectorial, los cuales optimizan la ejecución del gradientel.

1. Para finalizar, muchas gracias por su atención.