

IMT2220 Semestre 2025-1

Tarea 1

Elwin van 't Wout

8 de abril de 2025

Introducción

Los algoritmos de aprendizaje automatizado supervisado se basan en la predicción de una etiqueta (*label*) a través de variables (*features*) relevantes. Algunos ejemplos de estos algoritmos son los bosques aleatorios y las redes neuronales. En todo caso, los modelos deben ser entrenados con una base de datos que incluye casos conocidos. Este proceso de entrenamiento optimiza los parámetros del modelo tal que el resultado se ajuste lo mejor posible a los datos ya rotulados. Es decir, se minimiza el error de la predicción, medido por la función de pérdida, con un método de optimización. Como ejemplo, los parámetros en regresión lineal son la pendiente en cada dimensión y el intercepto de la línea recta. Además, la función de pérdida es la suma de diferencias cuadradas.

Se puede escribir el entrenamiento como

$$\min_{\mathbf{x} \in \mathbb{R}^m} F(\mathbf{x})$$

en lo cual F es la función de pérdida y \mathbf{x} los m parámetros del modelo. En general, la función de pérdida es una función no lineal y la cantidad de parámetros es grande. Por tanto, es casi nunca posible encontrar el mínimo como una expresión analítica. En su lugar, se aproxima el mínimo con un proceso iterativo $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_L$ en lo cual \mathbf{a}_i aproxima \mathbf{x} . Se elige la aproximación inicial \mathbf{a}_0 y se diseña un algoritmo tal que \mathbf{a}_j es una aproximación cada vez mejor del mínimo, hasta terminar en iteración L con la aproximación final \mathbf{a}_L .

Uno de los métodos iterativos de optimización más conocido es el *descenso de gradiente* (inglés: *gradient descent*). Este algoritmo busca el mínimo haciendo pasos en la dirección del descenso más grande. Más precisa, esta dirección es el vector gradiente negativo: $-\nabla F(\mathbf{a})$. Por lo tanto, el método usa la secuencia

$$\mathbf{a}_{j+1} = \mathbf{a}_j - \gamma \nabla F(\mathbf{a})$$

para $\gamma \in \mathbb{R}$ un parámetro llamado *tasa de aprendizaje* (inglés: *learning rate*). Bajo algunas condiciones, se puede demostrar que la secuencia converge a un mínimo local de la función F . Sin embargo, también existen casos en los que el método diverge.

Tarea

Esta tarea contempla la implementación del descenso de gradiente.

1. Para cada una de las funciones siguientes, calcule (a mano) el vector gradiente, todos los puntos críticos y el mínimo global. Escribe el resultado en el notebook de entrega.

$$f(x, y) = x^2 + y^2$$

$$g(x, y) = x^2 - y^2$$

$$h(x, y) = x^4 - x^3 - 2x^2 + |y|$$

2. Dibuje la gráfica de cada función.
3. Dibuje (en otra figura) varias curvas de nivel para cada función.
4. Dibuje todos los puntos críticos que calcularon para cada función como un *marker* en la figura anterior.
5. Implemente el algoritmo de descenso de gradiente.
6. Ejecute el algoritmo para los casos siguientes:

$$a) f(x, y); \gamma = 0,01; \mathbf{a}_0 = (2; -1)$$

$$b) f(x, y); \gamma = 1; \mathbf{a}_0 = (2; -1)$$

$$c) f(x, y); \gamma = 1,01; \mathbf{a}_0 = (2; -1)$$

$$d) g(x, y); \gamma = 0,01; \mathbf{a}_0 = (1; 0)$$

- e) $g(x, y)$; $\gamma = 0,01$; $\mathbf{a}_0 = (1; 0,01)$
 f) $h(x, y)$; $\gamma = 0,1$; $\mathbf{a}_0 = (-0,01; 0,5)$

En todos los casos, utilice suficientes iteraciones para ver la convergencia del algoritmo.

7. Dibuje toda la secuencia de aproximaciones, junto a varias curvas de nivel.
8. Por último, aplique el algoritmo a la función

$$k(x, y) = x^4 - 7x^3 - 14x^2 - 8x + y^2$$

sin calcular los puntos críticos. Use $\mathbf{a}_0 = (-0,1; 1)$ como punto inicial y $\gamma_1 = 0,14$, $\gamma_2 = 0,15$, $\gamma_3 = 0,16$ como distintas tasas.

Para cada combinación de función, tasa y punto inicial en las tareas, explique claramente el comportamiento del algoritmo. En específica: ¿por qué el algoritmo converge a este punto?

Evaluación

Entregue todo el código y las respuestas a las preguntas en un Jupyter notebook a través de Canvas.

Los reglamentos del curso se puede encontrar en Canvas. Se destaca que las tareas deben ser hechas de forma individual. No se puede compartir código entre compañeros, tampoco usar código de fuentes externos salvo el código proporcionado en Canvas. Se puede usar herramientas de inteligencia artificial (p.ej., ChatGPT y GitHub CoPilot) para la programación, pero las respuestas a las preguntas en las celdas de *markdown* deben ser escritos por ustedes mismos.